Characterization of genomic diversity in *cpn*60 defined *Enterococcus* ecotypes

Submitted to the College of Graduate Studies and Research of the University of Saskatchewan in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Veterinary Microbiology at the University of Saskatchewan.

By

ISHA KATYAL

PERMISSION TO USE

In agreement with the outlines set out by the College of Graduate Studies and Research at the University of Saskatchewan, I allow the University of Saskatchewan Libraries to make this thesis available to all interested parties. Also in accordance with the College of Graduate Studies and Research, I allow this thesis to be copied "in any manner, in whole or in part, for scholarly purposes". This thesis may not, however, be reproduced or used in any manner for financial gain without my written consent. Any scholarly use of this thesis, in part or in whole, must acknowledge both myself and the University of Saskatchewan.

Any requests for copying or using this thesis, in any form or capacity, should be made to:

Head of Department of Veterinary Microbiology

University of Saskatchewan

Saskatoon, Saskatchewan

S7N 5B4

ABSTRACT

The astounding complexity of microbial communities limits the ability to study the role of genomic diversity in shaping the community composition at the species level. With the advancement and increased affordability of high-throughput sequencing methods, it is increasingly recognized that genomic diversity at the sub-species level plays an important role in selection during microbial community succession.

Recent studies using the *cpn*60 universal target (UT) have shown that it is a high-resolution tool that provides superior resolution in comparison to 16S rRNA based tools and can predict genome relatedness. However, studies to characterize the nature and degree of genome content differences predicted by *cpn*60 UT sequence relationships have not been conducted. In this work, we focused on the *Enterococcus* community obtained from feces of healthy pigs. Enterococci are both accessible with selective culture, and interesting since the genus includes members that are significant human pathogens and others that are used as probiotics. Previous work has shown that *cpn*60 UT sequences can resolve pig fecal *Enterococcus faecalis* and *E. hirae* into phylogenetically and phenotypically distinct ecotypes.

The utility of *cpn*60 UT sequences for resolution of *Enterococcus* species was first established in the year 2000, and this demonstration included 17 *Enterococcus* species. We have expanded the analysis to include all currently recognized *Enterococcus* species and confirmed that *cpn*60 UT sequences provide higher resolution than 16S rRNA sequences for identification of *Enterococcus* species. In addition, we showed that

*cpn*60 UT sequences could resolve sub-groups in *E. faecium* consistent with results obtained from comparison of whole genome sequences.

GTG rep PCR based clusters for *E. faecalis* and *E. hirae* isolates were generally consistent with the *cpn*60 defined *Enterococcus* ecotypes observed in our previous study, suggesting that *cpn*60 UT sequences predict overall genomic relationships. Results from analysis of CRISPR sequences provided insights into the extensive networking and transfer of genetic material that takes place within the intestinal *Enterococcus* community. We conducted whole genome sequencing of representative isolates to characterize further the genomic diversity in *cpn*60 defined *E. hirae* ecotypes to determine the nature and degree of genome content differences predicted by *cpn*60 UT sequences. Differences in phosphotransferase systems, amino acid metabolism pathways for glutamine, proline and selenocystiene, potassium-transporting ATPases, copper homeostasis systems and putative prophage associated sequences, CRISPRs and antibiotic resistance genes were observed. Results from *in vitro* growth competition assays showed that isolates from *E. hirae-1* and *E. hirae-2* were able to out-compete isolates from the *E. hirae-3* ecotype, consistent with the relatively low abundance of *E. hirae-3* relative to *E. hirae-1* and *E. hirae-2* previously observed in the pig fecal microbiome, and with observed gene content differences between the ecotypes.

Results presented in this thesis provide a genomic basis for the definition of ecotypes within *E. hirae* and confirm the utility of the *cpn*60 UT sequence for high resolution profiling of complex microbial communities.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Janet Hill for her patience, motivation, immense knowledge and the continuous support during my PhD study. Her guidance helped me all the time in research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD study. I would like to thank the rest of my thesis committee: Dr. Peter Howard, Dr. Manuel Chirino, and Dr. Tim Dumonceaux, for their insightful comments and encouragement throughout my PhD.

My sincere thanks also goes to Dr. Bonnie Chaban and Champika Fernando. Without they precious support it would not have been possible to conduct this research.

I thank my fellow labmates - Atul, Matt, Teenus, John, Aline, Matheus, Jason and Lisa for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last few years. Also, my office mates – Rui, Iran, Yanyun, Shrijana, Arinjay and all other officemates – who made me feel at home with their kindness and patiently listened to all the daycare stories related to my son. Thank you Lana, Linda and Pat.

I would like to thank my family: my parents and to my brother and sisters for supporting me throughout writing this thesis and in my life in general.

Last but not the least, I would like to thank my husband who is a strong pillar of support for me and has loved me through all ups and downs during the PhD program. And to my son Aarya who motivates me to keep exploring and not be afraid of falling down.

Finally, I dedicate this thesis to my father-in-law who passed away too soon. He would have been the happiest person on earth at this day.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1 - Introduction and literature review

1.1 Succession in microbial communities

Succession, defined as a systematic and predictable manner by which biological communities change over time following a disruption in an existing environment or colonization of a new environment, is a universal phenomenon for all biological communities. Microbial communities are highly diverse, dynamic and comprise multiple "units" traditionally classified as species. The diversity in microbial communities, measured as the number of prokaryotic species and their relative abundance, is influenced by several factors such as competition for space and nutrition between community members, evolutionary aspects such as rate of accumulation of mutations, and environmental factors such as temperature and pH (Torsvik et al. 2002). To date, microbial succession has primarily been studied and understood using concepts established for plant and animal ecosystems and most of the ecological concepts about microbial community assembly have been derived from plant ecosystems (Fierer et al. 2010; Reynolds et al. 2003). Evidence accumulated in recent years has shown that microbial communities do not follow the ecological patterns that are well established for plant and animal communities. For example, Fierer et al. (2011) studied the relationship between elevation and bacterial diversity using 16S rRNA based sequencing methods and showed that elevation does not have a significant influence on bacterial diversity in contrast to patterns observed for plant and animal communities. Similarly, Bryant et al. (2008) showed that the patterns of diversity in plant and microbial communities with increasing elevation were distinct. One of the major reasons for lack of information

regarding microbial diversity and succession is the phylogenetic and phenotypic complexity of microbial communities, their astounding abundance in nature and the short generation time of microbes. There is a significant gap in our understanding of succession patterns in microbial communities and studies for the development of comprehensive concepts and models for microbial community succession are needed.

1.2 Methods used to study bacterial diversity and succession

1.2.1 Culture based methods

Initial studies of microbial diversity and succession relied on obtaining pure cultures of individual species that constitute bacterial communities followed by identification based on phenotypic characteristics (Winslow 1914). Techniques such as scanning electron microscopy, culture on selective media and biochemical tests to measure enzymatic activity over a period of time have also been used to study microbial succession (Federle et al. 1982; Harrison 1978; Marcellino and Benson 1992). Culture based methods provide phenotypic information for single isolates and significant information for taxonomic classification (Palleroni 1997) and in recent years, developments in culture based methods have resulted in isolation of previously uncultured bacteria from diverse environments and identification of novel antibiotics such as teixobactin (Aoi et al. 2009; Bollmann et al. 2007; Connon and Giovannoni 2002; Ferrari et al. 2005; Ling et al. 2015). However, culture based methods are time consuming and labour intensive and often fail to represent the actual diversity in complex microbial communities (Wagner et al. 1993). Discrepancies in viable bacterial counts obtained from culture plates and numbers of micro-organisms observed *via* microscopy

have been reported (Staley and Konopka 1985). Further, isolation of organisms from an environment generally fails to reproduce the ecological niches and the synergistic-antagonistic interactions that exist in the original microbial community.

1.2.2 Culture independent methods

The limitations of culture-based methods have been addressed with the use of molecular techniques that rely on the fact that cellular molecules can reflect evolutionary histories of organisms (Woese and Fox 1977). The use of molecular methods revolutionized the field of microbial ecology by revealing the astounding complexity of natural microbial communities and providing information about previously uncultured or undetected microbes (Hugenholtz et al. 1998; Hugenholtz and Pace 1996). Revolutionizing works of Torsvik et al. (1990) and Dykhuizen (1998) estimated that small amounts of pristine forest soils contain on the order of 10,000 species (with numbers up to 10^7) based on the rates of re-annealing of total bacterial DNA content.

1.2.2.1 16S rRNA based methods

Historically, one of the most commonly used molecular methods for characterizing diversity and studying succession in bacterial communities was the PCR amplification, cloning and sequencing of the 16S rRNA gene. The 16S rRNA gene is an approximately 1.5 kb gene that encodes a structural RNA and is universally conserved in bacteria and archeaea (Brosius et al. 1978). Giovannoni et al. (1990) published the first study of bacterial diversity using sequence libraries of cloned 16S rRNA genes. Their examination of samples from the Sargasso Sea demonstrated that culture based methods

failed to describe the diversity present in this environment. In a similar study by Ward et al. (1990), it was shown that previously unknown bacteria that were not detected using culture based methods could be detected using PCR, cloning and sequencing of 16S rRNA gene sequences. To date, a majority of the studies to characterize microbial community dynamics in environments such as the oceans, soil and human/animal-associated communities have used the 16S rRNA gene (Costello et al. 2009; Nemergut et al. 2011; Sogin et al. 2006), although cloning and Sanger sequencing has been replaced by higher throughput next-generation sequencing approaches.

In 16S rRNA based studies, total DNA from microbial communities is extracted and the 16S rRNA gene is amplified by PCR using "universal" primers. The resulting PCR product is then sequenced and compared to reference sequence databases for taxonomic and phylogenetic assignment. The sequence data is then binned into "operational taxonomic units" (OTU) that represent clusters of nearly identical sequences. The sequence micro-heterogeneity in these clusters is frequently ignored as PCR artifact owing to the use of non-proofreading polymerases (Acinas et al. 2005). Field et al. (1997) analyzed the 16S rRNA sequence micro-heterogeneity in the SAR 11 cluster within the bacterioplankton community of the Sargasso sea (Britschgi and Giovannoni 1991; Giovannoni et al. 1990) and proposed that sequence micro-heterogeneity could correspond to niche specialization and represent ecologically distinct populations within the microbial community. Further support for this idea came from 16S rRNA based denaturing gradient gel electrophoresis (DGGE) and sequence analyses of diversity in the bacterioplankton community of Lake Vilar over winter, spring, and summer, which showed that closely related populations of *Synechococcus* had different

phenotypic properties (pigment compositions) that co-related with differences in temporal distributions of these populations (Casamayor et al. 2002).

Even though 16S rRNA based studies have improved our understanding of microbial community composition and dynamics, knowing only a single gene's sequence provides limited information about an organism's role in a microbial community. Harris et al. (2004) reported that >70,000 rRNA sequences in publically available databases correspond to micro-organisms that have not been cultured and thus, phenotypic and genomic characteristics that determine the ecological functions of these organisms remain unknown. Further, Fox et al. (1992) showed that 16S rRNA sequences do not always correctly identify species, specifically for recently diverged species. Moreover, bacterial genomes are "flexible" in terms of the their genomic content and can alter their phenotypic characteristics (ecological functions) by frequent acquisition and loss of genes (Gogarten et al. 2002).

1.2.2.2 Protein-coding genes for microbial community profiling

Use of conserved protein coding genes for identification of ecologically distinct populations has been suggested as an alternative to 16S rRNA gene sequencing (Palys et al. 2000; Palys et al. 1997). Two protein-coding genes have been described for use in microbial profiling: *rpo*B (Mollet et al. 1997) and *cpn*60 (Goh et al. 1996).

1.2.2.2.1 RNA polymerase subunit beta (rpoB) gene

*rpo*B is a universally conserved gene that encodes the RNA polymerase beta subunit, which is required for RNA synthesis. It is a single copy gene (Case et al. 2007)

that has been proposed as a superior alternative to 16S rRNA for analysis of microbial community composition using denaturing gradient gel electrophoresis (DGGE) (Dahllof et al. 2000). It has extensively utilized for identification of bacterial species (Adekambi et al. 2009; Adekambi et al. 2008; Glazunova et al. 2009; Gundi et al. 2009). Vos et al. (2012) used *rpo*B based pyrosequencing methods for characterization of microbial diversity in soil and concluded that *rpo*B can complement 16S rRNA based studies. However, *rpo*B is limited by the unavailability of universal PCR primers and a reference sequence database.

1.2.2.2.2 60 kDa chaperonin *(cpn*60*)* gene

The *cpn*60 gene (also known as *gro*EL or *hsp*60) encodes the type I molecular chaperone, a 60 kDa protein that is involved in the folding and stability of proteins (Hemmingsen et al. 1988), is associated with the cell surface in some bacteria and can be involved in adherence of bacteria to gastric epithelial cells (Hennequin et al. 2001; Yamaguchi et al. 1999; Yamaguchi et al. 1997). It is found in bacteria and some archaea, and in the plastids, mitochondria, and cytoplasm of eukaryotes. A universal PCR system for amplification of a 549-567 bp region of the *cpn*60 gene corresponding to nucleotides 274-828 of the *E. coli cpn*60 gene (the "universal target", UT) has been established (Goh et al. 1996). The UT sequences are short enough that they can be sequenced in a single reaction using di-deoxy sequencing methods and provide better phylogenetic resolution than full-length 16S rRNA sequences which are >1kb in length.

The *cpn*60 UT has been used for species level identification for several genera including *Streptococcus*, *Enterococcus*, *Staphylococcus* and *Campylobacter*

(Dumonceaux et al. 2006a; Goh et al. 1998; Goh et al. 2000; Goh et al. 1997; Hill et al. 2006a; Hill et al. 2006b). Further, *cpn*60 provides a higher level of resolution at the species and sub-species levels when compared to 16S rRNA and other targets (Brousseau et al. 2001; Goh et al. 2000; Paramel Jayaprakash et al. 2012; Verbeke et al. 2011; Vermette et al. 2010). The universal PCR system for *cpn*60 has been optimized for detection of high G+C content organisms such as the Actinobacteria that are often under-represented in 16S rRNA microbial community studies (Hill et al. 2006b). A reference database, cpnDB, that includes *cpn*60 sequences from a broad range of taxa (approximately 20,000 sequences available at the time of writing) is available (www.cpndb.ca)(Hill et al. 2004). Availability of cpnDB allows for species specific primer design for use in detection and quantification of bacteria in environmental samples and for diagnostic purposes (Chaban et al. 2009; Chaban et al. 2010; Dumonceaux et al. 2009; Dumonceaux et al. 2011). Recently, the *cpn*60 UT has been established as a preferred barcode for bacteria (Links et al. 2012) and a predictor of genome relatedness in bacteria (Verbeke et al. 2011).

*cpn*60 UT based methods have provided comprehensive information about the diversity in a wide range of microbial communities including pulp and paper sludge (Dumonceaux et al. 2006b), the intestinal/fecal communities of cats, chickens and pigs and human fecal and vaginal communities (Chaban et al. 2013; Desai et al. 2009; Dumonceaux et al. 2006a; Hill et al. 2010; Hill et al. 2005; Hill et al. 2002). *cpn*60 UT has also been applied in pyrosequencing based studies for characterization of bacterial diversity in the human vaginal microbiome (Dumonceaux et al. 2009; Schellenberg et al. 2009; Schellenberg et al. 2011a; Schellenberg et al. 2011b), gastrointestinal tract of

rainbow trout (Desai et al. 2012), the human respiratory tract and the fecal microbial community of dogs (Chaban and Hill 2012; Chaban et al. 2012).

1.2.2.3 Fingerprinting methods

In addition to their use as targets for sequencing, PCR amplified products can be used for a variety of microbial community fingerprinting methods such as automated ribosomal intergenic spacer analysis (Fisher and Triplett 1999; Ranjard et al. 2006), terminal restriction fragment length polymorphism (Dunbar et al. 2000; Pesaro et al. 2004), denaturing gradient gel electrophoresis (Dahllof et al. 2000), single strand conformation polymorphism (Lee et al. 1996), and denaturing high-performance liquid chromatography (Priha et al. 2013). These methods have been demonstrated to be useful in providing information about change in bacterial community composition based on change in banding patterns of the community fingerprints.

1.2.3 Whole community meta-genomics

Comprehensive genomic data from the whole community can provide insights into genomic diversity that plays a critical role in defining microbial population structure, function, and succession, as well as revealing novel genes. A recent approach to understanding and characterizing diversity in bacterial communities is whole community shotgun sequencing where total community DNA is sequenced. This approach was successfully applied by Venter et al. (2004) who generated over one billion base pairs of non-redundant sequence data from the bacterioplankton community of the Sargasso Sea. At least 1800 genomic species were resolved from the sequence data and over 1.2 million

previously unknown genes were detected. A similar approach was used by Tyson et al. (2004) to characterize genomic heterogeneity in a biofilm community from an acid mine drainage. However, the whole community shotgun approach often fails to provide adequate sequence depth to assemble individual genomes and the data may be contaminated with host sequences when it is derived from environments like the gastrointestinal tract (Gevers et al. 2012; Kuczynski et al. 2012). Moreover, sequence data and assembly is computationally demanding (Nagarajan and Pop 2013).

1.3 Ecological units of bacterial diversity

Closely related strains within a species exhibit diversity in gene sequences and genome content. Strain-specific sequences may encode functions such as nutrient utilization and homeostasis mechanisms that are related to environmental niche specialization in an environment, and selection acting on these genes drives evolutionary diversification between strains (Jordan et al. 2002; Wu et al.).

The concept of ecotype was initially proposed for plants to describe genetically distinct varieties of plant species that are adapted to specific environmental conditions. A bacterial ecotype is defined as "*a phylogenetic group of closely related bacteria that are ecologically similar, in that the members of an ecotype share genetic adaptations to a particular set of habitats, resources, and conditions*" (Cohan 2002b). Further, ecotypes emerge in populations through periodic selection (Cohan 2002a, b). PCR based analysis of universal protein coding genes has allowed for identification of ecotypes within *Bacillus* (Connor et al. 2010; Koeppel et al. 2008), Vibrionaceae strains coexisting in a coastal bacterioplankton community (Hunt et al. 2008), *Leptospirillum* (Denef et al. 2010) and the marine cyanobacterium *Prochlorococcus* (Kettler et al. 2007). However,

little effort has been put toward characterization of the extent and type of genomic diversity that defines ecotypes within species. In a recent review by Shapiro and Polz (2014), a "reverse ecology" approach was suggested to study microbial community dynamics. In this approach, closely related isolates are chosen from a microbial community to represent ecotypes. These isolates can be chosen by the use of informative UTs that provide sub-species resolution. Phenotypic tests such as sole carbon source utilization among other biochemical tests should be performed to obtain insights into the functional capabilities of these ecotypes. Whole genome shotgun sequencing should be used to characterize genomic diversity in the ecotypes and to relate this diversity to observed differences in phenotypic characteristics. Core and pan-genome analysis should be conducted to identify unique genes in each ecotype. If the unique genes confer ecological advantage to the ecotypes, then it can be accepted that the ecotypes are ecologically and genetically distinct. This approach was applied to identify single nucleotide polymorphisms in previously published genomes of *Vibrio cyclitrophicus* (Shapiro et al. 2012) and *Sulfolobus islandicus* (Cadillo-Quiroz et al. 2012) to understand the role of genetic diversity in emergence of ecotypes. A similar approach was used by Kopac et al. (2014) where four closely related strains of *Bacillus subtilis* and laboratory strain W23 were compared to understand speciation in this model organism.

1.4 Microbial diversity in the gastrointestinal tract

Microbial diversity and succession in the gastrointestinal tract (GI) has been extensively explored due to its role as a determinant of animal and human health. The mammalian GI tract microbial community is complex, dynamic, and spatially diversified that comprises at least 10^{13} microorganisms (Luckey 1972) corresponding to > 800

species of bacteria (Kelly and Mulder 2011). The collective microbiota of the gut may contain up to 3 million genes (Qin et al. 2010). The study of microbial ecology of the GI tract includes identification of the gut microbes, their abundance and diversity, function (using biochemical tests and next generation sequencing methods such as transcriptomics, and metabolomics), and definition of the relationships of community members with each other, with the host and diet. Major advances in our knowledge about GI microbial communities can be attributed to the development of anaerobic culture techniques, and experimental animal model systems including gnotobiotic technology where germ-free animals can be colonized with a defined microbiota (Savage 2001).

In the last several decades, the role of the intestinal commensal microbiota in a wide range of host physiological functions such as development of the immune system (Mazmanian et al. 2005; Rakoff-Nahoum et al. 2004) and metabolic activities such as transport of water in colon, secretion of volatile fatty acids (Yolton and Savage 1976) and production of vitamin K (Ramotar et al. 1984) has been studied extensively (Inoue et al. 2005; Palmer et al. 2007). The gut microbial community structure is influenced by external environmental factors such as composition of maternal microbiota, diet, health and nutritional status of the animal. In addition, intrinsic or host-related factors such as intestinal pH, peristalsis, bile acids, host secretions, immune responses and bacterial mucosal receptors affect gut microbial diversity and succession (Inoue et al. 2005).

Both culture dependent and culture independent studies have shown that gut microbial community composition is not static but changes over time with host maturation and associated changes in diet, eventually resulting in a stable “climax community” in adults (Hohwy et al. 2001; Simpson et al. 2003). Studies focused on

characterizing changes in GI microbial community composition from birth to adulthood have provided insights into factors such as dietary changes and use of antimicrobials that may determine succession patterns in gut microbial communities (Berrington et al. 2013; Costello et al. 2009; Favier et al. 2002; Round and Mazmanian 2009; Sekirov et al. 2010; Torok et al. 2011). Studies focused on defining and predicting succession patterns in gut microbial communities are significant for utilization of microbes in applied research fields such as the development of alternatives to antibiotic drugs, alternate treatments for chronic gastrointestinal disorders and obesity, and production of probiotic foods.

1.4.1 Swine gastrointestinal microbial community

The pig gastrointestinal tract has been an attractive model system to study the role of gut microbiota as a determinant of animal health due to its close physiological resemblance to human gastrointestinal tract (Heinritz et al. 2013). Moreover, understanding of microbial diversity in pig feces is significant due to agricultural, public health and environmental concerns. Pig feces are easily available *via* non-invasive methods and have been studied for decades using both culture dependent and culture independent methods (Butine and Leedle 1989; Hill et al. 2002; Leser et al. 2002; Pryde et al. 1999; Robinson et al. 1981). Development of pig intestinal microbiota from the birth to the climax community of the adult has been well described (Isaacson and Kim 2012; Leser et al. 2002; Pedersen et al. 2013). Culture based methods have shown that aerobes and facultative anaerobes dominate the colon microbiome of piglets from the time of birth to the end of the first week of life and that this population is gradually replaced by anaerobes as the animal ages (Swords et al. 1993). Culture based methods

have shown that anaerobes, specifically those belonging to the genera *Bacteroides*, *Selenomonas*, and *Butyrivibrio* constitute the majority of fecal microbiota of adult pigs (Robinson et al. 1981).

To date, the majority of culture independent studies of the pig GI microbiome have utilized the 16S rRNA gene. One of the first studies describing the microbial composition of the swine GI tract using molecular methods was done by Pryde et al. (1999) who showed that 16S rRNA based sequencing methods could detect previously unknown bacterial phyla. However, this study along with other earlier investigations were limited to single time points and included small numbers of clones and/or sequences examined and hence, were not adequate for documenting and studying succession in the GI microbiome. Hill et al. (2002) showed that *cpn*60 UT sequences provided a higher resolution description of the pig fecal microbiome as compared to 16S rRNA based methods, but this study was also limited to a single time point in development. Kim et al. (2011) published a detailed analysis of microbial succession from 20 commercial pigs using high throughput pyrosequencing of 16S rRNA sequence libraries. Fecal samples were obtained from pigs at 10 weeks of age and then at three-week intervals until the pigs were 22 weeks old. Over the period of 12 weeks, a shift in community composition reflecting an overall increase in the proportion of *Firmicutes* and decrease in the proportion of *Bacteroidetes* was observed. However, the change in community composition was described only at the phylum and genus level due to limited resolving power of the 16S rRNA gene target.

The astounding complexity of the swine gut microbiome has limited the study of succession at the species level, and reliance on single gene targets has precluded

assessment of the genomic diversity that may play a role in shaping the community composition. Complementation of these whole community approaches with focused studies of micro-organisms that are easy to cultivate in laboratory settings offers an opportunity to expand our understanding of microbial succession at the species and subspecies level. For example, in a previous study, we identified *Enterococcus* isolates from pig feces using *cpn*60 UT sequencing and showed that a shift in the *Enterococcus* community composition occurs with change in diet over time (Vermette et al. 2010). The resolving power of the *cpn*60 UT sequence revealed succession occurring not only at the species level, but also at the sub-species level. However, the genomic diversity among these isolates and the relationship of genome content to phenotype and observed succession pattern was not explored in this study.

1.5 The genus *Enterococcus* – history and classification

Enterococci were first described as Gram-positive, spherical, intestinal micro-organisms and termed “entérocoque” by Thiercelin (1899). In another publication by Maccallum and Hastings (1899), an *Enterococcus* like organism was identified from a patient with endocarditis and classified as *Micrococcus zymogenes*; an organism that was later renamed *Enterococcus faecalis.* These authors described the organism as "very hardy and tenacious of life" as is true for clinical *E. faecalis* and *E. faecium* isolates. Andrewes and Horder (1906) used the name *Streptococcus faecalis* to describe an organism obtained from a patient with endocarditis that was phenotypically similar to the organism described by Maccallum and Hastings (1899). The first description of *Streptococcus faecium* was provided by Orla-Jensen (1919) as morphologically similar to *Streptococcus faecalis* but had different fermentation capabilities from *Streptococcus*

faecalis. Sherman and Wing (1935) described *Streptococcus hemothermophilus* as a non-pathogenic, beta-hemolytic, Gram positive organism that does not ferment sucrose. *Streptococcus hemothermophilus* was later re-classified as *E. durans*.

The genus *Enterococcus* was classified as group D streptococci (Sherman 1937) until Schleifer and Kilpper-Balz (1984) showed that *S. faecalis* and *S. faecium* were different from other members of the genus *Streptococcus* using DNA-DNA hybridization. Collins et al. (1984) showed that *Streptococcus avium*, *Streptococcus casseliflavus*, *Streptococcus durans* and *Streptococcus gallinarum,* were closely related to the genus *Enterococcus* based on biochemical and genetic properties. The newly established genus *Enterococcus* was defined as belonging to the family Enterococcaceae, order Lactobacillales, Class Bacilli and division Firmicutes (Ruan 2013).

1.5.1 Genus features

The genus *Enterococcus* comprises Gram positive facultative anaerobes, ovoid in shape that appear in single cells, in pairs, or in short chains on Gram staining (Thiercelin 1899). They are catalase negative, non-spore forming and have the ability to survive in extreme environments with high pH, temperature and salt concentrations (Bliss 1937; Sherman 1937; Thiercelin 1899). They have the ability to grow in broth containing 6.5% NaCl, can hydrolyze esculin in the presence of 40% bile salts and are usually homofermentative (Klein 2003; Murray 1990; Ruan 2013). The G + C content of DNA ranges from 37 to 45 mol%. They are lactic acid producers and produce bacteriocins (Franz et al. 2007; Nes et al. 2007). Species in this genus, with the exception *E. gallinarum* and *E. casseliflavus*, are non-motile (Collins et al. 1986; Graudal 1957). *E.*

sulfureus, *E. casseliflavus*, and *E. mundtii* are usually associated with plants, and colonies of these species appear yellow due to the presence of a pigment in the cell wall (Collins et al. 1986; Graudal 1957; Martinez-Murcia and Collins 1991).

Enterococcus species are commonly isolated from the intestines of animals (Baele et al. 2002; Devriese et al. 1992a; Devriese et al. 1994; Devriese et al. 1991; Devriese et al. 1992c), food products (Ben Omar et al. 2004; Franz et al. 2003; Mannu et al. 1999; Suzzi et al. 2000; Yousif et al. 2005), insects (Martin and Mundt 1972) and the environment (Mundt 1961; Svec and Sedlacek 1999). They can be a serious issue in human medicine because of the presence of antibiotic resistance genes, virulence factors such as the enterococcal surface protein (ESP), cytolysin and aggregation substance, and their frequent association with hospital-acquired infections. On the other hand, they occur naturally in dairy products (Mannu and Paba 2002; Mannu et al. 1999; Suzzi et al. 2000) and a few strains have been shown to have probiotic effects in pigs and humans (Broom et al. 2006; Macha et al. 2004; Szabo et al. 2009). Host specific *Enterococcus* strains have been identified in the gastroinsteinal tracts of animals (Gilmore 2002). *E. faecalis*, *E. faecium*, *E. hirae*, *E. mundtii* and *E. avium* are reported among the most commonly occurring *Enterococcus* species in the pig intestine (Facklam et al. 2002).

1.5.2 Methods of identification

Several systems have been developed for rapid identification of enterococcal isolates based on their biochemical/phenotypic features (Facklam et al. 2002). One of these is a test using esculin in a buffered NaCl solution, in which enterococci produce a black color due to hydrolysis of esculin. Some other systems use a combination of Gram

stain, hydrolysis of L-pyrrolidonyl-β-napthylamide and reaction to group D and group A streptococcal anti-sera. Differentiation of the two most commonly isolated enterococci, *E. faecalis* and *E. faecium*, is done by exploiting their biochemical differences such as the ability of *E. faecalis* but not *E. faecium* to grow on a medium containing 0.04% tellurite and to reduce tetrazolium to formazan. Selective media such as mEnterococcus agar (Slanetz and Bartley 1957) are available for one-step detection and isolation of *Enterococcus* species. Commercially available kits for phenotypic identification of *Enterococcus* isolates (mainly, *E. faecalis* and *E. faecium*) are available and include API 20S, API Rapid ID 32 and the Vitek system among others (Facklam et al. 2002).

Enterococcus species have been classified into five phenotypic groups (I-V) based on hydrolysis of arginine, motility, pigment production, pyruvate utilization, tolerance to 0.04% tellurite, hippurate hydrolysis and acid formation in broth containing arabinose, glycerol, raffinose, sucrose, trehalose, xylose, methyl-α-D-glucopyranoside, mannitol and sorbose (Facklam and Collins 1989; Facklam et al. 2002). Group I consists of enterococcal species that form acid in carbohydrate broths, but do not hydrolyze arginine and includes *E. phoeniculicoa, E. devriesei, E. canis, E. avium, E. malodoratus, E. raffinosus, E. pseudoavium, E. saccharolyticus, E. pallens,* and *E. gilvus*. Group II includes *E. casseliflavus, E. mundtii, E. gallinarum, E. canintestini, E. lactis, E. thailandicus,* and *E. sanguinicola* and mannitol positive variants of *E. faecalis* and *E. faecium* species that form acid in mannitol broth and hydrolyze arginine but fail to form acid in sorbose broth. Group III includes species that hydrolyze arginine but are not able to form acid in either mannitol or sorbitol broth. This group comprises *E. villorum, E. durans, E. dispar, E. hirae, E. silesiacus,* and *E. rotai,* as well as mannitol negative

variants of *E. faecalis* and *E. faecium. E. aquamarinus, E. plantarum, E. caccae, E. termitis, E. asini, E. sulfureus* and *E. cecorum* belong to Group IV. These species neither form acid in mannitol and sorbose broth nor hydrolyze arginine. Group V includes *E. columbae, E. rivorum, E. hermaniensis, E. camelliae,* and *E. viikiensis,* as well as variants of *E. casseliflavus, E. gallinarum,* and *E. faecalis* that form acid on mannitol, but not sorbose broth, and do not have the ability to hydrolyse arginine. Group VI includes *E. ureilyticus* that can form acid from sorbose but not in mannitol broth and does not hydrolyze arginine. *E. quebecensis, E. italicus,* and *E. ureasiticus* have not yet been classified into this scheme due to lack of availability of fermentation data.

The development of molecular methods has resulted in improved knowledge about diversity in this genus. Whole cell protein analysis has been shown to distinguish up to 21 *Enterococcus* species (Niemi et al. 1993; Teixeira et al. 2001; Teixeira et al. 1997a). 16S rRNA sequences are available for all recognized *Enterococcus* species and phylogenetic analysis has been conducted (Lebreton et al. 2014). However, 16S rRNA sequences cannot resolve all the *Enterococcus* species (Cotta et al. 2013; Patel et al. 1998). Randomly amplified polymorphic DNA (RAPD) analysis has been shown to be useful for accurate identification of clinical *Enterococcus* isolates (Quednau et al. 1998). Other molecular methods used for identification and characterization of isolates from the genus *Enterococcus* include sequencing of the domain V of the 23S rRNA gene (Naimi et al. 1999; Tsiodras et al. 2000), amplification of rRNA or tRNA intergenic spacers (Tyrrell et al. 1997), sequencing of the D-ala:D-ala ligase genes (*ddl*) (Ozawa et al. 2000), the manganese-dependent superoxide dismutase (*sodA*) genes (Frolkova et al. 2012; Poyart et al. 2000), sequencing and hybridization of the *cpn*60 UT (Goh et al.

2000) and application of multilocus sequence analysis that targets the RNA polymerase α-subunit gene (*rpoA*) and the phenylalanyl-tRNA synthase gene (*pheS*) (Naser et al. 2005a). Multi locus sequence typing (MLST) has been used for characterizing genomic diversity in *E. faecalis* (McBride et al. 2007) and *E. faecium* isolates (Homan et al. 2002; Ruiz-Garbajosa et al. 2006). Genomic fingerprinting techniques such pulsed-field gel electrophoresis (PFGE) and PCR amplification of repetitive extragenic palindromic PCR (REP-PCR), BOX-PCR and GTG rep PCR have been used for identification and study of diversity in commensal and clinical *Enterococcus* isolates (De Vuyst et al. 2008; Gordillo et al. 1993; Malathum et al. 1998; Nayak et al. 2011; Papaparaskevas et al. 2002; Svec et al. 2005c; Teixeira et al. 1997a).

1.5.3 *Enterococcus* genomics

Draft genomes for over 700 enterococcal strains have been published at the time of writing (http://www.ncbi.nlm.nih.gov/genome). In addition, a large-scale effort to sequence more than 500 *Enterococcus* genomes including clinical, animal and healthy human isolates was initiated by the Broad Institute (*Enterococcus* Illumina PacBio Initiative, Broad Institute (broadinstitute.org)). The size of genomes in this genus ranges from 2.5 to 3.5 Mb. A considerable effort has been put towards comparative genomics of *E. faecalis* and *E. faecium* as they are associated with the majority of hospital acquired infections (Willems et al. 2005). The genome of *E. faecalis* strain V583, a clinical isolate, was the first complete *Enterococcus* genome published (Bourgogne et al. 2008; Paulsen et al. 2003).

The genus *Enterococcus* has been shown to exhibit genome plasticity (Aakra et al. 2007; Lepage et al. 2006). About 25% of genomes of strains from this genus comprise insertion sequences (IS) elements, phages, and genetic elements of various kinds that include virulence factors and plasmids. The relative contribution of these elements may vary from strain to strain but all species probably contain representatives of each type. Mobile elements in enterococci can be generally classified into three different systems: conjugative broad host plasmids, conjugative pheromone responsive plasmids and conjugative transposons. These mobile elements have been well characterized in *E. faecalis* and *E. faecium*, the two most clinically relevant species. Genome sequences for other *Enterococcus* species have become available more recently. Draft genome sequences are now available for strains of *E. caccae, E. asini, E. avium, E. caccae, E. casseliflavus, E. cecorum, E. columbae, E dispar, E. durans, E. gallinarum, E. gilvus, E. haemoperoxidus, E. italicus, E. malodoratus, E. moraviensis, E. mundtii, E. pallens, E. phoeniculicola, E. raffinosus, E. saccharolyticus, E. sulfureus,* and *E. villorum* (Magni et al. 2012; Palmer et al. 2010; Palmer et al. 2012). Availability of sequence data from genomes of other *Enterococcus* species will improve our understanding of the genomic diversity in this genus.

Results of a comparison of the genomes of two *E. faecalis* strains (V583 and OG1RF) demonstrated that environmental factors such as the origin of the strain are associated with genome content differences (Bourgogne et al. 2008). To date, comparative genomic studies have focused on pathogenic *E. faecalis* and *E. faecium* strains due to their association with hospital-acquired infections. However, most of this diverse genus comprises non-pathogenic *Enterococcus* species occupying a variety of

environments such as the GI tract, plant surfaces, food and soil. To date, these strains remain comparatively unexamined at the genome level. Although a large scale-sequencing project to understand genomic diversity in this genus has been initiated by the Broad Institute, no comparative studies have been published for species other than *E. faecalis* and *E. faecium*. Comparative genomic studies for species in this genus will provide information on genetic factors that play a role in enterococcal community dynamics, and niche specialization.

1.5.4 *Enterococcus hirae*

Enterococcus hirae has been used as a model organism for basic research for over four decades owing to its ability to grow in laboratory conditions (Gaechter et al. 2012). Mechanism of copper homeostasis in Gram positive bacteria have been explored extensively in *E. hirae* (Lu et al. 2003; Odermatt et al. 1993; Pecou et al. 2006; Wunderli-Ye and Solioz 1999). Further, it has been used as a model organism to study bioenergetics of phosphate, potassium, sodium, and calcium transport (Bakker and Harold 1980; Burkler and Solioz 1982; Harold et al. 1965; Heefner and Harold 1982; Kobayashi et al. 1978; Murata et al. 1996; Odermatt et al. 1993; Solioz and Davies 1994). The *E. hirae* type strain (ATCC 9790) was first isolated from chickens (Farrow and Collins 1985). *E. hirae* is found in soil, is a commensal in the gastrointestinal tract of humans, insects, and farm animals such as pigs and chickens, and is recognized as a causative agent in endocarditis in humans and animals (Canalejo et al. 2008; Devriese et al. 1992b; Devriese and Haesebrouck 1991; Gilad et al. 1998). The complete genome sequence of *E. hirae* ATCC 9790 (Gaechter et al. 2012), and draft genome sequences of

three *E. hirae* strains from the human gut and cultured milk have been published recently (strains: INF E1, SB10 and EnGen0127). However, comparative genomic studies among *E. hirae* isolates or comparisons with other *Enterococcus* species have not been published to date.

This research project focuses on the *Enterococcus* community in swine feces. Even though enterococci are not numerically dominant members of adult pig intestinal microbiota, their ease of isolation, dual nature (pathogenic and commensal), genome plasticity and ability to utilize a wide variety of carbohydrates makes them a good model for studying intestinal community dynamics at the genome level. In a previous study (Vermette et al. 2010), we showed that *cpn*60 UT sequences were able to resolve subgroups within *E. faecalis* (*E. faecalis-1*, *E. faecalis-2* and *E. faecalis-3*) and *E. hirae* (*E. hirae-1*, *E. hirae-2*, *E. hirae-3* and *E. hirae-4*) that were not apparent using 16S rRNA sequences. Further, a shift in the *Enterococcus* community composition occurs with change in diet over time, suggesting that selection (due to different diets and age of the pig) plays an important role in community development and succession. Each subgroup likely corresponds to an ecotype (a set of strains using the same ecological niche), but the consistency of *cpn*60-based clustering and whole genome relationships between these groups has not been established. Confirmation of ecotype level resolution within natural microbial communities would be a major contribution to justification for use of the *cpn*60 UT sequence as the preferred tool for microbial community profiling.

OBJECTIVES

The overall objectives of this project are to:

1. Determine if *cpn*60 UT sequences can resolve species and sub-species in the genus *Enterococcus,* including new species defined since 2000.
2. Identify phenotypic and genotypic features that differentiate *cpn*60 defined subgroups of *Enterococcus hirae* (proposed ecotypes) and relate these to the observed succession in the pig fecal *Enterococcus* community.

CHAPTER 2 - The *cpn*60 universal target: A useful tool to resolve enterococci

Second Chapter Transition

In this chapter, the utility of the *cpn*60 UT to differentiate all *Enterococcus* species known to date has been described. This was determined by analyzing publically available *cpn*60 UT sequences and comparing the results to the widely used techniques such as the 16S rRNA based sequencing methods, whole genome sequence comparisons and fingerprinting techniques such as MLST. The results presented in this chapter show that *cpn*60 UT provides higher resolution power to differentiate *Enterococcus* species and sub-groups within *E. faecium* in comparison to 16S rRNA sequencing methods.

2.1 Abstract

Nomenclature of species within the genus *Enterococcus* is continuously being revised. Previous work by our group showed that the *cpn60* UT is a useful tool for identification of enterococci, but since the time of that study, the number of species in the genus *Enterococcus* has increased to from 17 to 54. The objective of this study was to re-visit the previous study and determine if *cpn*60 UT sequences can resolve species and sub-species groups in the genus *Enterococcus*. We determined the phylogenetic relationships between *cpn*60 UT sequences of all the enterococcal species for which data is publically available (n=28) including *cpn*60 UT sequences from larger collections of *E. faecium* (n=156) and *E. faecalis* (n=182) isolates. Our results show that *cpn*60 UT sequences can distinguish enterococcal species and furthermore, can resolve *E. faecium* strains into two distinct clades, which correspond with the results from multilocus sequence typing (MLST) and comparative genomics study of *E. faecium*. Phylogenetically distinct clusters were not seen among *E. faecalis* isolates suggesting that it is a more homogenous taxon in comparison to *E. faecium*. In conclusion, we show that *cpn*60 remains a useful tool to identify *Enterococcus* species and also has potential utility in sub-species resolution.

2.2 Introduction

Enterococcus spp. are Gram positive facultative anaerobes, ovoid in shape, and appearing in single cells, in pairs, or in short chains. General characteristics of this genus include their ability to grow at 10 to 45 ºC, in 6.5% NaCl, and to hydrolyze esculin in the presence of bile. They are commonly isolated from intestines of animals (Baele et al. 2002; Devriese et al. 1992a; Devriese et al. 1994; Devriese et al. 1991; Devriese et al. 1992c), food products (Ben Omar et al. 2004; Franz et al. 2003; Mannu et al. 1999; Suzzi et al. 2000; Yousif et al. 2005), insects (Martin and Mundt 1972) and the environment (Mundt 1961; Svec and Sedlacek 1999). Several methods have been developed for rapid identification of *Enterococcus* species based on their biochemical properties (Devriese et al. 1993; Facklam and Elliott 1995; Manero and Blanch 1999). However, the utility of these biochemical tests is limited by different biochemical profiles revealed by strains of the same species isolated from different sources (Jackson et al. 2004). Further, these tests have been validated for only a limited number of *Enterococcus* species, limiting their use for identification of unusual or novel species.

With advances in molecular biology techniques, sequence-based identification of *Enterococcus* species and sub-species has been investigated. To date, *Enterococcus faecalis* and *E. faecium*, the two species for which the most sequence data has been generated, have been classified using multiple methods such as 16S rRNA gene sequencing (Williams et al. 1991), whole genome sequencing (Palmer et al. 2012), and MLST (Willems et al. 2005). 16S rRNA gene sequence based phylogenetic analysis has provided limited resolution for distinguishing *Enterococcus* species. For example, the 16S rRNA gene sequences for *E. casseliflavus* and *E. gallinarum* are 99.9% identical,

erroneously suggesting that they are the same species (97% identity is the threshold typically used for differentiating species) (Stackebrandt and Goebel 1994). MLST is similar to 16S rRNA gene sequencing as both techniques rely upon the sequencing of conserved genomic loci, but MLST provides subspecies resolution. MLST can be a reliable tool for molecular epidemiology of *E. faecium* isolates but is not appropriate for species identification as it is more time consuming and expensive than a single target PCR. At the time of writing, the MLST scheme for *E. faecalis* includes six housekeeping genes: glucose-6-phosphate dehydrogenase (*gdh*), glyceraldehydes-3-phosphate dehydrogenase (*gyd*), phosphate ATP binding cassette transporter (*pstS*), glucokinase (*gki*), shikimate-5-dehydrogenase (*aroE*), xanthine phosphoribosyltransferase (*xpt*), and acetyl-CoA-acetyltransferase (*yiqL*) (http://efaecalis.mlst.net/misc/info.asp). Similarly, the MLST scheme for *E. faecium* includes PCR for seven housekeeping genes: adenosine kinase (*adk*), ATP synthase (*atpA*), D-alanyl-alanine synthetase (*ddl*), glucose-6-phosphate dehydrogenase (*gdh*), glyceraldehydes-3-phosphate dehydrogenase (*gyd*), carboxyaminoimidazole ribonucleotide synthase (*purK*) and phosphate ABC transporter (*pstS*) (http://efaecium.mlst.net/misc/info.asp). Several nucleic acid fingerprinting methods including rep-PCR (Svec et al. 2005c), randomly amplified polymorphic DNA (RAPD) fingerprinting (Quednau et al. 1998), and PFGE have been used for *Enterococcus* species identification (Pangallo et al. 2008), and studying genomic diversity of *Enterococcus* species from clinical and environmental sources.

The *cpn*60 UT is a region of the gene encoding the 60 kDa chaperonin found in bacteria, eukaryotes and some archaea, corresponding to nucleotides 274-828 of the *E. coli cpn*60 gene (Hill et al. 2004). The *cpn*60 UT has been shown to be a useful tool for

identification of microorganisms (Blaiotta et al. 2008; Brousseau et al. 2001; Minana-Galbis et al. 2009; Sakamoto and Ohkuma 2010) including two recently recognized *Enterococcus* species: *Enterococcus lemanii* sp. nov. and *Enterococcus eurekensis* (Cotta et al. 2013). It is a single target that can be amplified with universal PCR primers, and provides more resolving power than the 16S rRNA gene, especially for closely related taxa. For example, *E. faecalis* and *E. hirae* isolates can be grouped into phylogenetically and phenotypically distinct ecotypes using *cpn*60 UT sequences but not 16S rRNA gene sequences (Vermette et al. 2010). The *cpn*60 UT has been established as a preferred barcode for bacteria (Links et al. 2012) and has been applied in gene-based metagenomic studies of complex microbial communities (Chaban et al. 2012; Hill et al. 2002; Schellenberg et al. 2009; Schellenberg et al. 2011a). Further, *cpn*60 UT based applications are supported by a reference database of chaperonin sequences, cpnDB (Hill et al. 2004).

Previous work by our research group has demonstrated that *cpn*60 UT sequences can resolve enterococcal species and distinguish them from related species of *Lactococcus* and *Vagococcus* (Goh et al. 2000). But since the time of that study (2000), the number of species in the genus *Enterococcus* has increased to from 17 to 54 (Parte 2014). The objective of the current study was to evaluate *cpn*60 UT sequences as a tool to discriminate species and resolve subspecies groups within enterococci. Availability of an accessible and useful tool to resolve species and sub-species and that can predict genome relatedness will have utility for diagnostics, as well as for application in research to understand the role of enterococci in microbial communities.

2.3 Material and Methods

2.3.1 Sequence sources

The list of all *Enterococcus* species validly published to date was obtained from the List of Prokaryotic Names with Standing in Nomenclature (LPSN, (Parte 2014)) and is provided in Table 2.1. 16S rRNA sequences for all species were obtained from NCBI Genbank. *cpn*60 UT sequences were obtained from the chaperonin database, cpnDB (www.cpndb.ca) (Hill et al. 2004). The 16S rRNA accession numbers and cpnDB IDs (where available) for all species are provided in Table 2.1.

Table 2.1 Currently recognized *Enterococcus* species

Species	Type strain	cpnDB ID*	16S rRNA accession	Reference
Enterococcus alcedinis	CCM 8433	NA	JX948102	Frolkova et al. (2013)
Enterococcus aquimarinus	CCUG 51308	b22921	AJ877015	Svec et al. (2005a)
Enterococcus asini	ATCC 700915	b26568	Y11621	de Vaux et al. (1998)
Enterococcus avium	ATCC 14025	b815	AF133535	Collins et al. (1984)
Enterococcus caccae	ATCC BAA-1240	b26566	AY943820	Carvalho Mda et al. (2006)
Enterococcus camelliae	NBRC 101868	NA	EF154454	Sukontasing et al. (2007)
Enterococcus canintestini	CCUG 3785	NA	AJ888906	Naser et al. (2005b)
Enterococcus canis	NBRC 100695	NA	X76177	de Graef et al. (2003)
Enterococcus casseliflavus	ATCC 25788	b819	AF039903	Collins et al. (1984)
Enterococcus cecorum	ATCC 43198	b822	AF061009	Williams et al. (1989)
Enterococcus columbae	ATCC 51263	b3380	AF061006	Devriese et al. (1990)
Enterococcus devriesei	CCUG 37865	NA	AJ891167	Svec et al. (2005b)
Enterococcus diestrammenae	KACC 16708	NA	JQ650245	Kim et al. (2013)
Enterococcus dispar	ATCC 51266	b26570	AF061007	Collins et al. (1991)
Enterococcus durans	ATCC 19432	b823	AJ276354	Collins et al. (1984)
Enterococcus eurekensis	CCUG 61259	b26675	AF445305	Cotta et al. (2013)
Enterococcus faecalis	ATCC 19433	b821	AB012212	Schleifer and Kilpper-Balz (1984)
Enterococcus faecium	ATCC 19434	b818	AJ301830	Schleifer and Kilpper-Balz (1984)
Enterococcus flavescens[1]	ATCC 49996	b7608	AJ420802	Pompei et al. (1992)
Enterococcus gallinarum	ATCC 49573	b825	AF039900	Collins et al. (1984)
Enterococcus gilvus	CCUG 45553	b7607	DQ411810	Tyrrell et al. (2002)
Enterococcus haemoperoxidus	ATCC BAA-382	b7606	AF286832	Svec et al. (2001)
Enterococcus hermanniensis	CCUG 48100	NA	AY396047	Koort et al. (2004)
Enterococcus hirae	ATCC 8043	b824	Y17302	Farrow and Collins (1985)
Enterococcus italicus	DSM 15952	b18294	AJ582753	Fortina et al. (2004)
Enterococcus lactis	LMG 25958	NA	GU983697	Morandi et al. (2012)
Enterococcus lemanii	CCUG 61260	b26679	AF445301	Cotta et al. (2013)
Enterococcus malodoratus	ATCC 43197	b817	AJ301835	Collins et al. (1984)
Enterococcus moraviensis	ATCC BAA-383	b7605	AF286831	Svec et al. (2001)

Species	Type strain	cpnDB ID*	16S rRNA accession	Reference
Enterococcus mundtii	ATCC 43186	b820	AF061013	Collins et al. (1986)
Enterococcus olivae	DSM 25431	NA	KJ566121	Lucena-Padros et al. (2014)
Enterococcus pallens	CCUG 45554	b7604	DQ411812	Tyrrell et al. (2002)
Enterococcus phoeniculicola	ATCC BAA-412	b7603	AY028437	Law-Brown and Meyers (2003)
Enterococcus plantarum	CCM 7889	NA	HQ847537	Svec et al. (2012)
Enterococcus porcinus	ATCC 700913	NA	AF335596	Teixeira et al. (2001)
Enterococcus pseudoavium	ATCC 49372	b814	AF061002	Collins et al. (1989b)
Enterococcus quebecensis	CCUG 59306	NA	GU457262	Sistek et al. (2012)
Enterococcus raffinosus	ATCC 49427	b816	Y18296	Collins et al. (1989a)
Enterococcus ratti	ATCC 700914	NA	AF539705	Teixeira et al. (2001)
Enterococcus rivorum	CCM 7986	NA	FN822765	Niemi et al. (2012)
Enterococcus rotai	CCUG 61593	NA	AJ276353	Sedlacek et al. (2013)
Enterococcus saccharolyticus	ATCC 43076	b813	AF061004	Rodrigues and Collins (1990)
Enterococcus saccharominimus[2]	CCUG 50447	NA	AJ626902	Vancanneyt et al. (2004)
Enterococcus seriolicida[3]	ATCC 49156	NA	L32813	Kusuda et al. (1991)
Enterococcus silesiacus	CCM 7319	NA	AM039966	Svec et al. (2006)
Enterococcus solitarius[4]	ATCC 49428	b7602	AJ301840	Collins et al. (1989b)
Enterococcus sulfureus	ATCC 49903	b7601	AF061001	Martinez-Murcia and Collins (1991)
Enterococcus termitis	CCM 7300	NA	AM039968	Svec et al. (2006)
Enterococcus thailandicus	NBRC 101867	NA	EF197994	Tanasupawat et al. (2008)
Enterococcus ureasiticus	CCUG 59304	NA	GU457264	Sistek et al. (2012)
Enterococcus ureilyticus	CCUG 48799	NA	AJ276352	Svec et al. (2006)
Enterococcus viikkiensis	LMG 26075	NA	HQ378515	Rahkila et al. (2011)
Enterococcus villorum[5]	CCUG 45025	b830	AJ271329	Vancanneyt et al. (2001)
Enterococcus xiangfangensis	NCIMB 14834	NA	HF679036	Li et al. (2014)

*NA = not available in publically available data sources
[1] Synonym of *Enterococcus casseliflavus*
[2] Synonym of *Enterococcus italicus*
[3] Renamed as *Lactococcus garvieae*
[4] Renamed as*Tetragenococcus solitarius*
[5] Synonym of *Enterococcus porcinus*

Genome sequences, either completed or in progress, of *E. faecium* and *E. faecalis* strains used in the study were downloaded from the Broad Institute's Olive database (https://olive.broadinstitute.org/projects/enterogenome). The *cpn*60 UT sequences of all strains were located within the genome sequences by using *E. faecalis* V583 (NCBI accession ID: NC_004668) as a query sequence and BLASTn to search the genomic sequences. Only full-length *cpn*60 UT sequences (552 bp) were included in the analysis. A total of 156 *E. faecium* and 182 *E. faecalis* strains were included and are listed in Table 2.2 and Table 2.3, respectively. Metadata for *E. faecium* strains related to the original isolation sources, countries of origin and isolation dates were obtained from a database of *E. faecium* MLST profiles (http://efaecium.mlst.net/), PATRIC database (Wattam et al. 2014) or the NCBI genome database (http://www.ncbi.nlm.nih.gov/genome).

Table 2.2 *Enterococcus faecium* strains included in the study

Strain	Sub-group	Source of isolation	Country[1]	Year of isolation	References
109a1	1	Human	Denmark	1995-1998	(Wattam et al. 2014)
1230933	1	Hospitalized patient (wound)	USA	2005	
1231408	1	Hospitalized patient (blood)	USA	2005	
1231410	1	Hospitalized patient (skin and soft tissue infection)	USA	2005	(Palmer et al. 2012)
1231501	1	Hospitalized patient (blood)	USA	2005	
1231502	1	Hospitalized patient (blood)	USA	2005	
7230532	1	Pig (feces)	Denmark	2000	
7330381-1	1	Pig (feces)	Denmark	2001	
7330446-2	1	Pig (feces)	Denmark	2001	
7330519-3	1	Pig (feces)	Denmark	2001	
7330614-1	1	Pig (feces)	Denmark	2001	
7330884-2	1	Pig (feces)	Denmark	2001	
7430166-3	1	Pig (feces)	Denmark	2001	
841v03	1	Human (feces)	Denmark	2003	
9730219-1	1	Pig (feces)	Denmark	1997	
9730357-1	1	Pig (feces)	Denmark	1997	
9731349-1	1	Pig (feces)	Denmark	1997	
9731352-4	1	Pig (feces)	Denmark	1997	
9830091-5	1	Pig (feces)	Denmark	1998	
9830512-2	1	Pig (feces)	Denmark	1998	
9830565-4	1	Pig (feces)	Denmark	1998	
9930238-2	1	Pig (feces)	Denmark	1999	
9931110-4	1	Pig (feces)	Denmark	1999	
a17sv1	1	Pig (feces)	Denmark	1995	(Wattam et al. 2014)
atcc8459	1	Cheese	Unknown	Unknown	
e0045	1	Chicken (feces)	UK	1992	
e0120	1	Hospitalized patient (ascites)	Netherlands	1995	
e0164	1	Turkey (feces)	Netherlands	1996	
e0269	1	Turkey (feces)	Netherlands	1996	
e0333	1	Hospitalized patient (blood)	Israel	1997	
e0679	1	Pig (feces)	Belgium	Unknown	
e0680	1	Pig (feces)	Germany	Unknown	
e0688	1	Pig (feces)	Spain	Unknown	
e1050	1	Non-hospitalized patient (feces)	Netherlands	1998	
e1133	1	Hospitalized patient (feces)	USA	2001	
e1185	1	Hospitalized patient	France	Unknown	
e1258	1	Hospitalized patient (blood)	Spain	Unknown	
e1293	1	Hospitalized patient (blood)	Italy	Unknown	
e1321	1	Hospitalized patient (catheter)	Italy	1999	
e1392	1	Hospitalized patient	UK	2000	
e1552	1	Hospitalized patient (feces)	Netherlands	2002	

Strain	Sub-group	Source of isolation	Country[1]	Year of isolation	References
e1573	1	Bison (rumen)	Belgium	1994	
e1574	1	Dog (feces)	Belgium	1995	
e1575	1	Chicken (feces)	Belgium	1995	
e1576	1	Ostrich (caecum)	South Africa	2001	
e1578	1	Pig (feces)	Germany	2001	
e1620	1	Hospitalized patient (blood)	Netherlands	1957	
e1622	1	Mouse	Netherlands	1959	
e1623	1	Hospitalized patient (pus)	Netherlands	1960	
e1626	1	Hospitalized patient (stomach)	Netherlands	1965	
e1627	1	Hospitalized patient (gut)	Netherlands	1979	
e1630	1	River water	Netherlands	1981	
e1634	1	River water	Netherlands	1982	
e1644	1	Hospitalized patient	Germany	2002	
e1731	1	Hospitalized patient (blood)	Tanzania	Unknown	
e1904	1	Hospitalized patient (urine)	Netherlands	2001	
e2071	1	Poultry	Denmark	2001	
e2134	1	Chicken	Netherlands	2004	
e2297	1	Hospitalized patient (urine)	USA	2001	
e2369	1	Hospitalized patient	Hungary	2005	
e2560	1	Hospitalized patient (blood)	Netherlands	2006	
e2883	1	Hospitalized patient (blood)	Netherlands	2002	
e2966	1	Hospitalized patient (blood)	Netherlands	2005	
e3346	1	Hospitalized patient (blood)	Netherlands	2002	
e4215	1	Chicken	Sweden	2004	
e4389	1	Dog (feces)	Denmark	Unknown	
e6012	1	Hospitalized patient	Latvia	2010	
e6045	1	Hospitalized patient	Portugal	2010	
e8sv3	1	Pig (feces)	Denmark	1995	
en0305	1	Unknown	France	1994	
en0308	1	Unknown	Switzerland	1994	
f9730129-1	1	Chicken	Denmark	1997	
h17243	1	Human	Denmark	1995	
h17494	1	Human	Denmark	1995	
h17575	1	Human	Denmark	1995-1998	
hf50104	1	Pig (feces)	USA	2008	
hf50105	1	Pig (feces)	USA	2008	
hf50106	1	Pig (feces)	USA	2008	
hf50203	1	Pig (feces)	USA	2008	
hf50204	1	Pig (feces)	USA	2008	
hf50215	1	Pig (feces)	USA	2008	
hm1071	1	Unknown	Unknown	1994	
hm1072	1	Unknown	Unknown	1994	

Strain	Sub-group	Source of isolation	Country[1]	Year of isolation	References
hm1073	1	Unknown	Unknown	1994	
hm1074	1	Unknown	Unknown	1994	
ny1-1	1	Unknown	Unknown	Unknown	
ny1-2	1	Unknown	Unknown	Unknown	
ny1-3	1	Unknown	Unknown	Unknown	
ny1-4	1	Unknown	Unknown	Unknown	
ny1-5	1	Unknown	Unknown	Unknown	
ny1-6	1	Unknown	Unknown	Unknown	
ny2-1	1	Unknown	Unknown	Unknown	
ny2-2	1	Unknown	Unknown	Unknown	
ny2-3	1	Unknown	Unknown	Unknown	
ny3-1	1	Unknown	Unknown	Unknown	
ny1-1	1	Unknown	Unknown	Unknown	
uaa1007	1	Unknown	Unknown	1996	
uaa1019	1	Unknown	Switzerland	1996	
uaa1022	1	Unknown	Switzerland	1996	
uaa1023	1	Unknown	France	1996	
uaa1024	1	Unknown	France	1996	
uaa1025	1	Unknown	France	1996	
uaa1433	1	Unknown	France	2000	
uaa1484	1	Unknown	Unknown	1992	
uaa210	1	Unknown	Unknown	1986	
pip820	1	Unknown	Unknown	Unknown	
uaa430	1	Unknown	France	1989	
uaa431	1	Unknown	France	1989	
uaa714	1	Unknown	France	1994	
uaa715	1	Unknown	France	1994	
uaa716	1	Unknown	France	1994	
uaa718	1	Unknown	France	1994	
uaa719	1	Unknown	France	1994	
uaa720	1	Unknown	France	1993	
uaa721	1	Unknown	France	1994	
uaa722	1	Unknown	France	1994	
uaa723	1	Unknown	France	1993	
uaa724	1	Unknown	France	1993	
uaa725	1	Unknown	France	1993	
uaa825	1	Unknown	France	1996	
uaa909	1	Unknown	Switzerland	1996	
uaa910	1	Unknown	Switzerland	1996	
uaa911	1	Unknown	Switzerland	1996	
uaa944	1	Unknown	USA	1996	
uaa945	1	Unknown	USA	1996	

Strain	Sub-group	Source of isolation	Country[1]	Year of isolation	References
uaa947	1	Unknown	USA	1996	
uaa949	1	Unknown	USA	1996	
uaa950	1	Unknown	USA	1996	
uaa951	1	Unknown	USA	1996	
uaa952	1	Unknown	USA	1996	
van219	1	Chicken (feces)	Denmark	2010	
van222	1	Chicken (feces)	Denmark	2010	
van327	1	Unknown	Denmark	2010	
van332	1	Chicken (feces)	Denmark	2010	
van335	1	Chicken (feces)	Denmark	2010	
van342	1	Chicken (feces)	Denmark	2010	
van345	1	Chicken (feces)	Denmark	2010	
van476	1	Chicken (feces)	Denmark	2010	
vre108	1	Human	Denmark	2010	
vre110	1	Human	Denmark	2010	
vre13	1	Human	Denmark	2010	
vre84	1	Human	Denmark	2010	
com12	2	Non-hospitalised person (feces)	USA	2006	(Palmer et al. 2012)
com15	2	Non-hospitalised person (feces)	USA	2006	
8.19E+03	2	Pig feces	Canada	2007	(Vermette et al. 2010)
uaa1280	2	Unknown	France	1998	
e3548	2	Hospitalized patient (blood)	Netherlands	2004	
e1590	2	Non-hospitalised person (feces)	Ireland	2001	
e2620	2	Hospitalized patient (blood)	Netherlands	2006	
e3083	2	Hospitalized patient (blood)	Netherlands	2000	
1141733	2	Hospitalized patient (blood)	USA	2005	(Wattam et al. 2014)
e2039	2	Hospitalized patient (K-Spitze)	Germany	2000	
e1972	2	Hospitalized patient (blood)	Germany	2000	
e1861	2	Hospitalized patient (feces)	Spain	2001	
e1613	2	Fish burger	Norway	1964	
e1604	2	Cheese	Norway	1956	
e1007	2	Non-hospitalised person (feces)	Netherlands	1998	

Unknown = not available in publically available databases and scientific publications

Table 2.3 *Enterococcus faecalis* strains included in the study

Strain	Source of isolation	Country[1]	Year of isolation	Reference
599951	Human (blood)	USA	1994	(McBride et al. 2007), (Wattam et al. 2014)
79-3	Human (blood)	USA	1999	
ATCC 4200	Human (blood)	Unknown	Unknown	
b653	Human (blood)	Unknown	1956	
ch570	Human (blood)	USA	1987	
merz151	Human (blood)	USA	2002	
merz192	Human (blood)	USA	2002	
merz204	Human (blood)	USA	2002	
merz89	Human (blood)	USA	2002	
merz96	Human (blood)	USA	2002	
mmh594	Human (blood)	USA	1985	
rm3817	Human (blood)	USA	1960	
rm4679	Human (blood)	USA	1960	
sf21520	Human (blood)	Spain	Mid 1990s	
sf21521	Human (blood)	Spain	1990	
tr161	Human (blood)	USA	1993	
tr197	Human (blood)	USA	1993	
v583	Human (blood)	USA	1987	
ss-7	Cheese	Unknown	1918	
ATCC 35038	Chicken (intestine)	Unknown	1980	(Wattam et al. 2014)
12107	Human (clinical)	USA	Mid 1990s	
5952	Human (clinical)	USA	1975	
ds16	Human (clinical)	USA	1978	
fa2-2	Human (clinical)	UK	1973	
hip11704	Human (clinical)	USA	2002	
jh1	Human (clinical)	UK	1974	
rc73	Human (clinical)	USA	1979	
rmc1	Human (clinical)	USA	1954	
rmc5	Human (clinical)	USA	1954	
sf100	Human (clinical)	USA	Mid 1980s	
sf105	Human (clinical)	USA	Mid 1980s	
sf1592	Human (clinical)	USA	Late 1980s	

Strain	Source of isolation	Country[1]	Year of isolation	Reference
sf19	Human (clinical)	USA	Late 1980s	
sf339	Human (clinical)	USA	1986	
sf350	Human (clinical)	Canada	1986	
sf370	Human (clinical)	USA	1986	
sf6375	Human (clinical)	USA	1991	
Com 1	Non-hospitalised person (feces)	USA	2006	
Com 2	Non-hospitalised person (feces)	USA	2006	
Com 6	Non-hospitalised person (feces)	USA	2006	
Com 7	Non-hospitalised person (feces)	USA	2006	
pan7	Non-hospitalised person (feces)	USA	2005	
ATCC 27959	Cow (bovine mastitis isolate)	USA	1975	
ar01 dg	Dog (wound)	New Zealand	2001	
fly1	Drosophila	USA	2005	
fly2	Drosophila	USA	2005	
e1	Hospitalized patient (endocarditis)	USA	1960s	
ch116	Human (feces)	USA	1987-1988	
e1sol	Non-hospitalised person (feces)	USA	1960s	
ned10	Horse	Netherlands	1961	
ch136	Human (urine)	USA	1987-1988	
hh22	Human (urine)	USA	1982	
sf24396	Human (urine)	USA	2001	
sf24397	Human (urine)	USA	2001	
sf24413	Human (urine)	USA	2002	
sf26630	Human (urine)	USA	2002	
sf5039	Human (urine)	USA	1991	
t10	Human (urine)	USA	1992	
t12	Human (urine)	USA	1992	
t13	Human (urine)	USA	1992	
t14	Human (urine)	USA	1992	
t2	Human (urine)	USA	1992	

Strain	Source of isolation	Country[1]	Year of isolation	Reference
t3	Human (urine)	USA	1992	
t4	Human (urine)	USA	1992	
t5	Human (urine)	USA	1992	(Wattam et al. 2014)
t6	Human (urine)	USA	1992	
t7	Human (urine)	USA	1992	
t8	Human (urine)	USA	1992	
t9	Human (urine)	USA	1992	
v587	Human (urine)	USA	1987	
wh245	Human (urine)	USA	1987	
wh257	Human (urine)	USA	1987	
wh571	Human (urine)	USA	1986	
t16	Infant (feces)	UK	1951	
t17	Infant (feces)	UK	1951	
t18	Infant (feces)	UK	1951	
t19	Infant (feces)	UK	1951	
t20	Infant (feces)	UK	1951	
t21	Infant (feces)	UK	1951	
x98	Infant (feces)	Unknown	1934	
yi6-1	Human (clinical)	Japan	1992	
a-2-1	Infant (sepsis)	USA	Early 1980s	
a-3-1	Infant (sepsis)	USA	Early 1980s	
b-4-111	Infant (sepsis)	USA	Early 1980s	
ch188	Human (liver)	USA	Late 1980s	
ATCC 6055	Milk	Unknown	1937	
f1	Milk	Unknown	Early 1900 s	(Wattam et al. 2014)
39-5	Human (periodontitis)	Unknown	1964	
d1	Pig	Denmark	Unknown	(McBride et al. 2007)
d3	Pig	Denmark	Unknown	
d6	Pig	Denmark	Unknown	
1448e03	Human (feces)	Denmark	2003	(Wattam et al. 2014)
182970	Human (blood)	Denmark	2003	
19116	Pig (pork meat)	Denmark	2002	
2630v05	Human (feces)	Denmark	2005	

Strain	Source of isolation	Country[1]	Year of isolation	Reference
2924	Turkey meat	Denmark	2005	
atcc_10100	Unknown	Unknown	1948	
atcc_19433	Unknown	Unknown	1942	
atcc_27275	Unknown	Unknown	1962	
b1005	Human (blood)	USA	1985	
b1138	Human (blood)	USA	1985	
b1249	Human (blood)	USA	1985	
b1290	Human (blood)	USA	1985	
b1327	Human (blood)	USA	1985	
b1376	Human (blood)	USA	1985	
b1385	Unknown	Unknown	Unknown	
b1441	Human (blood)	USA	1986	
b1505	Human (blood)	USA	1986	
b1532	Human (blood)	USA	1986	(Wattam et al. 2014)
b15725	Hospitalized patient	Denmark	2007	
b1586	Human (blood)	USA	1986	
b1618	Human (blood)	USA	1986	
b1623	Human (blood)	USA	1986	
b16457	Hospitalized patient	USA	2007	
b1678	Human (blood)	USA	1986	
b1696	Human (blood)	USA	1986	
b1719	Human (blood)	USA	1986	
b1734	Human (blood)	USA	1986	
b1843	Human (blood)	USA	1986	
b1851	Human (blood)	USA	1986	
b1874	Human (blood)	USA	1986	
b1921	Human (blood)	USA	1986	
b1933	Human (blood)	USA	1986	
b2202	Human (blood)	USA	1986	
b2207	Human (blood)	USA	1986	
b2211	Human (blood)	USA	1986	
b2255	Human (blood)	USA	1986	
b2277	Human (blood)	USA	1986	

Strain	Source of isolation	Country[1]	Year of isolation	Reference
b2391	Human (blood)	USA	1986	
b2488	Human (blood)	USA	1986	
b2535	Human (blood)	USA	1986	
b2557	Human (blood)	USA	1986	
b2593	Human (blood)	USA	1986	
b2670	Human (blood)	USA	1986	
b2685	Human (blood)	USA	1986	
b2687	Human (blood)	USA	1986	
b2802	Human (blood)	USA	1987	
b2813	Human (blood)	USA	1987	
b2864	Human (blood)	USA	1987	
b2867	Human (blood)	USA	1987	
b2949	Human (blood)	USA	1987	
b3031	Human (blood)	USA	1987	
b3042	Human (blood)	USA	1987	
b3053	Human (blood)	USA	1987	
b3119	Human (blood)	USA	1987	
b3126	Human (blood)	USA	1987	
b3196	Human (blood)	USA	1987	
b3286	Human (blood)	USA	1987	
b3336	Human (blood)	USA	1987	
b4008	Human (blood)	USA	1987	
b4018	Human (blood)	USA	1987	
b4148	Human (blood)	USA	1987	
b4163	Human (blood)	USA	1987	(Wattam et al. 2014)
b4259	Human (blood)	USA	1987	
b4267	Human (blood)	USA	1987	
b4270	Human (blood)	USA	1987	
b4411	Human (blood)	USA	1987	
b4568	Human (blood)	USA	1987	
b4638	Human (blood)	USA	1987	
b4672	Human (blood)	USA	1987	
b4674	Human (blood)	USA	1987	

Strain	Source of isolation	Country[1]	Year of isolation	Reference
b4969	Human (blood)	USA	1987	
b5035	Human (blood)	USA	1988	
b5076	Human (blood)	USA	1987	
b56765	Hospitalized patient	Denmark	2007	
b594	Human (blood)	USA	1985	
b69486	Hospitalized patient (aortic vascular prosthesis)	Denmark	2007	
b84847	Hospitalized patient (hepato-biliary-pancreatic)	Denmark	2007	
b878	Human (blood)	USA	1985	(McBride et al. 2007)
b939	Human (blood)	USA	1985	
c_19315_led_1a_wt	Hospitalized patient (amyloid arthropaty)	Denmark	2004	
c_19315_led_1b_pp_scv	Hospitalized patient (amyloid arthropaty)	Denmark	2004	
ds5	Hospitalized patient	Unknown	NA	
e99	Unknown	Unknown	2000	
hef39	Human (blood)	Denmark	2002	
rmc65	Unknown	USA	1963	
sf28073	Human (urine)	USA	2003	
ss-6	Unknown	Unknown	1930	
t1	Unknown	Unknown	1950	
atcc_29200	Human (urogenital)	Canada	1974	
ch19	Human (wound)	USA	1987	
t15	Human (wound)	USA	1973	

Unknown = not available in publically available

2.3.2 Phylogenetic analysis

16S rRNA (n=54) and *cpn*60 UT sequences (n=28, available at the time of writing) of type strains of *Enterococcus* species and additional strains of *E. faecium* (n=156) and *E. faecalis* (n=182) were used to construct phylogenetic trees. Sequences were aligned with ClustalW (Thompson et al. 1994) using the default parameters (gap open penalty=10, gap extension penalty=0.20, gap distances=5). For 16S rRNA sequences, the resulting multiple sequence alignment was trimmed to 1273 shared nucleotides using Genedoc (Nicholas and Nicholas 1997). The Phylip software package (Felsenstein 1993) was used to calculate distance matrices using *protdist* for amino acid sequence alignments and *dnadist* for DNA sequence alignments, respectively. Alignments were bootstrapped using *seqboot* and trees were built using *neighbor* within Phylip, followed by consensus tree calculation using *consense*. The resulting consensus trees were visualized using Figtree (http://tree.bio.ed.ac.uk/software/figtree/). Pairwise nucleotide identities were calculated using SIAS (http://imed.med.ucm.es/Tools/sias.html) and pairwise distances for all the 156 *E. faecium* strains were calculated using MEGA version 6 (Tamura et al. 2013).

2.4 Results and Discussion

2.4.1 Identification of 16S rRNA and *cpn*60 UT sequences for *Enterococcus* spp.

Presently, a total of 54 different species have been proposed in the genus *Enterococcus* (as listed in the LPSN database) (Parte 2014), including species with a

wide range of phenotypic properties, isolated from diverse habitats (Table 2.1). Species identified since the year 2000 have been isolated from animal hosts, plants, soil, water and fermented food and dairy products. Out of all the known *Enterococcus* species, *E. faecalis* and *E. faecium* account for the most human enterococcal infections (Agudelo Higuita and Huycke 2014). Other species, including *E. avium*, *E. gallinarum*, *E. casseliflavus*, *E. hirae*, *E. mundtii*, and *E. raffinosus*, *E. sanguinicola*, *E. gilvus*, *E. pallens* and *E. canintestini* have also been associated with human infections, although less frequently (Carvalho Mda et al. 2008; Devriese et al. 1995; Hammerum 2012; Murray 1990; Tan et al. 2010; Tyrrell et al. 2002).

The 54 species names that have been proposed apparently correspond to only 49 unique species of *Enterococcus*. *E. saccharominimus* (Vancanneyt et al. 2004) has been proposed as a synonym of *E. italicus* (Fortina et al. 2004; Naser et al. 2006) and *E. casseliflavus* is an earlier synonym of *E. flavescens* (Naser et al. 2006; Teixeira et al. 1997b; Teixeira et al. 1996). *Enterococcus porcinus* (Teixeira et al. 2001) was shown to be the same species as *Enterococcus villorum* (Vancanneyt et al. 2001), as they are phenotypically and genotypically identical (de Graef et al. 2003). Additionally, *Enterococcus seriolicida* was reclassified as *Lactococcus garvieae* based on comparison of whole cell genomic and proteomic profiles with other isolates of *Lactococcus garvieae* (Elliott et al. 1991; Kusuda et al. 1991). *Enterococcus solitarius* was renamed as *Tetragenococcus solitarius* due to its close resemblance to other strains from the genus *Tetragenococcus* (Collins et al. 1989a; Ennahar and Cai 2005). Interestingly, the nucleotide and amino acid *cpn*60 UT sequences for *E. casseliflavus* and *E. flavescens*, *E. italicus* and *E. saccharominimus*, and *E. villorum* and *E. porcinus*, are identical. 16S

rRNA and *cpn*60 UT sequences from only one of each synonym pair are included in the phylogenetic analysis. Further, sequences for *E. seriolicida* (re-classified as *Lactococcus garvieae*) and *E. solitarius* (re-classified as *Tetragenococcus solitaries*) have not been included in the phylogenetic analysis.

16S rRNA sequences were available for all 49 *Enterococcus* species, but the *cpn*60 UT sequences of only 28 out of 49 *Enterococcus* species were available. For the remaining 21 *Enterococcus* species, sequences of other housekeeping genes such as the RNA polymerase α-subunit (*rpoA*), the phenylalanyl-tRNA synthase (*pheS*) and the elongation factor Tu (*tufA*) have been used to define their relationship to other *Enterococcus* species (Naser et al. 2005a; Picard et al. 2004).

The *cpn*60 UT sequences of 156 *E. faecium* and 182 *E. faecalis* strains were located within genome sequences as described in the Methods. Only one copy of the *cpn*60 UT was present in each genome sequence and all were 552 bp in length, which is consistent with the length of *cpn*60 UT sequence for low G+C Firmicutes.

2.4.2 Comparison of 16S rRNA and *cpn*60 UT based resolution of *Enterococcus* spp.

16S rRNA sequences for 49 *Enterococcus* species were retrieved from NCBI Genbank using the accession numbers provided by the LPSN (Table 2.1). A phylogenetic tree of 16S rRNA sequences is shown in Figure 2.1, which illustrates that 16S rRNA sequences can resolve only 32 of the 49 species. The 16S rRNA sequences of *E. pseudoavium, E. viikkiensis* and *E. devriesei*, and *E. gilvus* and *E. raffinosus* are identical. Further, the 16S rRNA sequences for *E. viikkiensis, E. devriesei, E pseudoavium, E. xiangfangensis, E. gilvus, E. gallinarium* and *E. raffinosus* are all more than 99% similar

to each other, failing to meet the generally accepted 97% cut-off value for species differentiation (Kim et al. 2014). Another example where lack of species resolution is evident is the *E. faecium* group that includes *E. faecium, E. lactis, E. durans, E. hirae, E. ratti, E. canis, E. villorum* and *E. mundtii* where 16S rRNA sequences for these species have pairwise identities in the range of 98.2–99%. The similarities for 16S rRNA sequences for species in the *E. faecalis* group that includes *E. rivorum, E. termitis, E. silesiacus, E. rotai, E. urealyticus, E. haemoperoxidus, E. moravensis, E. quebecensis, E. plantarum, E. caccae* and *E. ureasiticus* are 97.5-99.8%. Overall, the pairwise 16S rRNA sequence similarities for all species range from 94-100%. This lack of resolution is not surprising as similar observations have been made earlier by Lebreton et al. (2014) where 16S rRNA sequences from 40 distinct *Enterococcus* species were examined, and Naser et al. (2005a) where use of multilocus sequence analysis (MLSA) based on a combination of two house-keeping genes (*rpoA* and *pheS*) was shown to provide higher resolution than the 16S rRNA sequencing for identification of closely related species.

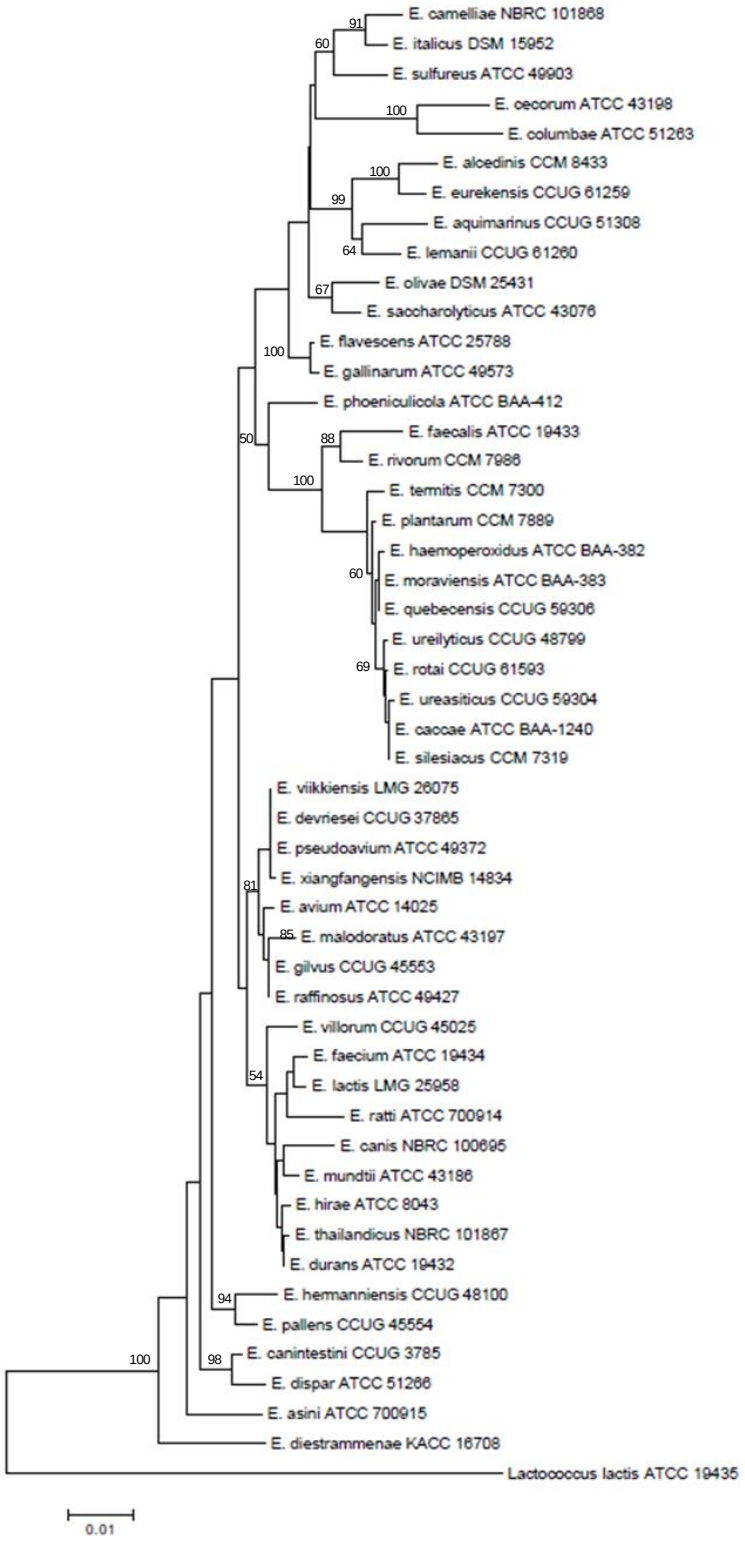

E. camelliae NBRC 101868
E. italicus DSM 15952
E. sulfureus ATCC 49903
E. cecorum ATCC 43198
E. columbae ATCC 51263
E. aloedinis CCM 8433
E. eurekensis CCUG 61259
E. aquimarinus CCUG 51308
E. lemanii CCUG 61260
E. olivae DSM 25431
E. saccharolyticus ATCC 43076
E. flavescens ATCC 25788
E. gallinarum ATCC 49573
E. phoeniculicola ATCC BAA-412
E. faecalis ATCC 19433
E. rivorum CCM 7986
E. termitis CCM 7300
E. plantarum CCM 7889
E. haemoperoxidus ATCC BAA-382
E. moraviensis ATCC BAA-383
E. quebecensis CCUG 59306
E. ureilyticus CCUG 48799
E. rotai CCUG 61593
E. ureasiticus CCUG 59304
E. caccae ATCC BAA-1240
E. silesiacus CCM 7319
E. viikkiensis LMG 26075
E. devriesei CCUG 37865
E. pseudoavium ATCC 49372
E. xiangfangensis NCIMB 14834
E. avium ATCC 14025
E. malodoratus ATCC 43197
E. gilvus CCUG 45553
E. raffinosus ATCC 49427
E. villorum CCUG 45025
E. faecium ATCC 19434
E. lactis LMG 25958
E. ratti ATCC 700914
E. canis NBRC 100695
E. mundtii ATCC 43186
E. hirae ATCC 8043
E. thailandicus NBRC 101867
E. durans ATCC 19432
E. hermanniensis CCUG 48100
E. pallens CCUG 45554
E. canintestini CCUG 3785
E. dispar ATCC 51266
E. asini ATCC 700915
E. diestrammenae KACC 16708
Lactococcus lactis ATCC 19435
91
60
100
100
99
64
67
100
50
88
100
60
69
81
85
54
94
100
98
0.01

Figure 2.1 Phylogenetic tree for the genus *Enterococcus* using the 16S rRNA sequences for type strains of 49 *Enterococcus* species.
The tree was constructed using the neighbor-joining method and is based on an alignment length of 1273 nucleotides. *Lactococcus lactis* is used as an outgroup. Bootstrap values were generated over 100 iterations. Bootstrap values >50 have been included in the figure.

Phylogenetic relationships of 28 *Enterococcus* species based on *cpn*60 UT sequences are shown in Figure 2.2. Sequences of all species included in the study were distinct, with pairwise sequence identities from 75-89%. To facilitate direct comparison of *cpn*60 UT and 16S rRNA phylogenies, the 16S rRNA based phylogenetic relationships of this subset of 28 *Enterococcus* species were determined (Figure 2.3). The 16S rRNA pairwise sequence similarities ranged from 94.4-100%. *cpn*60 UT based groups correlate well with the 16S rRNA sequence based groupings with the exception of the species in the *E. faecalis* (*E. phoeniculicols*, *E. moraviensis*, *E. haemoperoxidus* and *E. caccae*) and *E. faecium* (*E. villorum*, *E. mundtii*, *E. durans* and *E. hirae*) groups that could not be resolved by 16S rRNA sequences. The pairwise percent similarities between *cpn*60 UT sequences for *E. gilvus* and *E. raffinosus*; *E. durans* and *E. hirae*; *E. moraviensis* and *E. haemoperoxidus* are 88%, 87.3% and 87.1%, respectively, compared to 100% for 16S rRNA sequences. These results clearly show that for these species, *cpn*60 UT sequences have more discriminatory power than 16S rRNA sequences. Similar results have been seen in other studies where resolution into species and sub-species is not observed using 16S rRNA sequencing (Schellenberg et al. 2009).

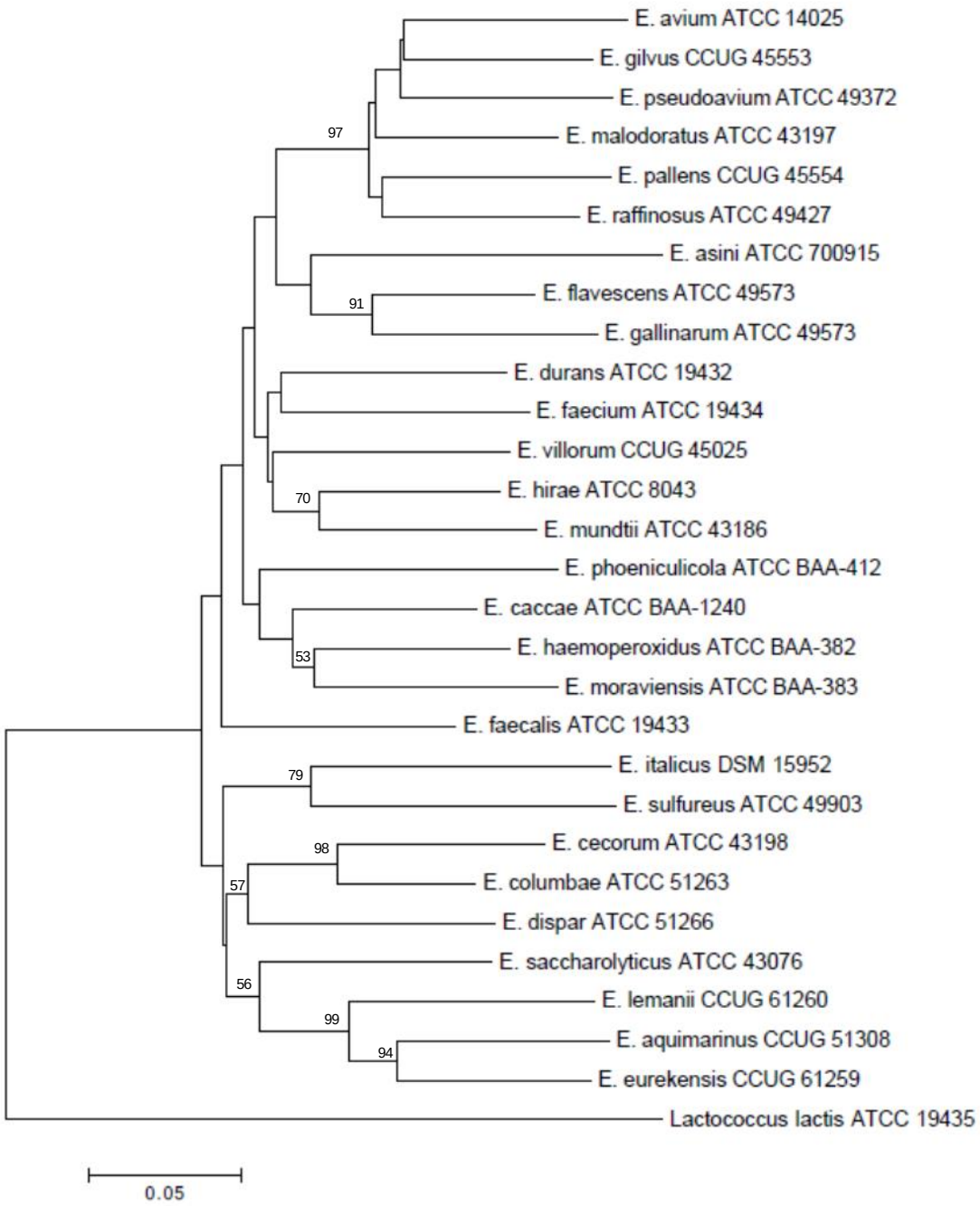


Figure 2.2 Phylogenetic tree of *cpn*60 UT nucleotide sequences for type strains of 28 *Enterococcus* species.

The tree was constructed using the neighbor-joining method and is based on an alignment length of 552 nucleotides. *Lactococcus lactis* is used as an outgroup. Bootstrap values were generated over 100 iterations. Bootstrap values >50 have been included in the figure.

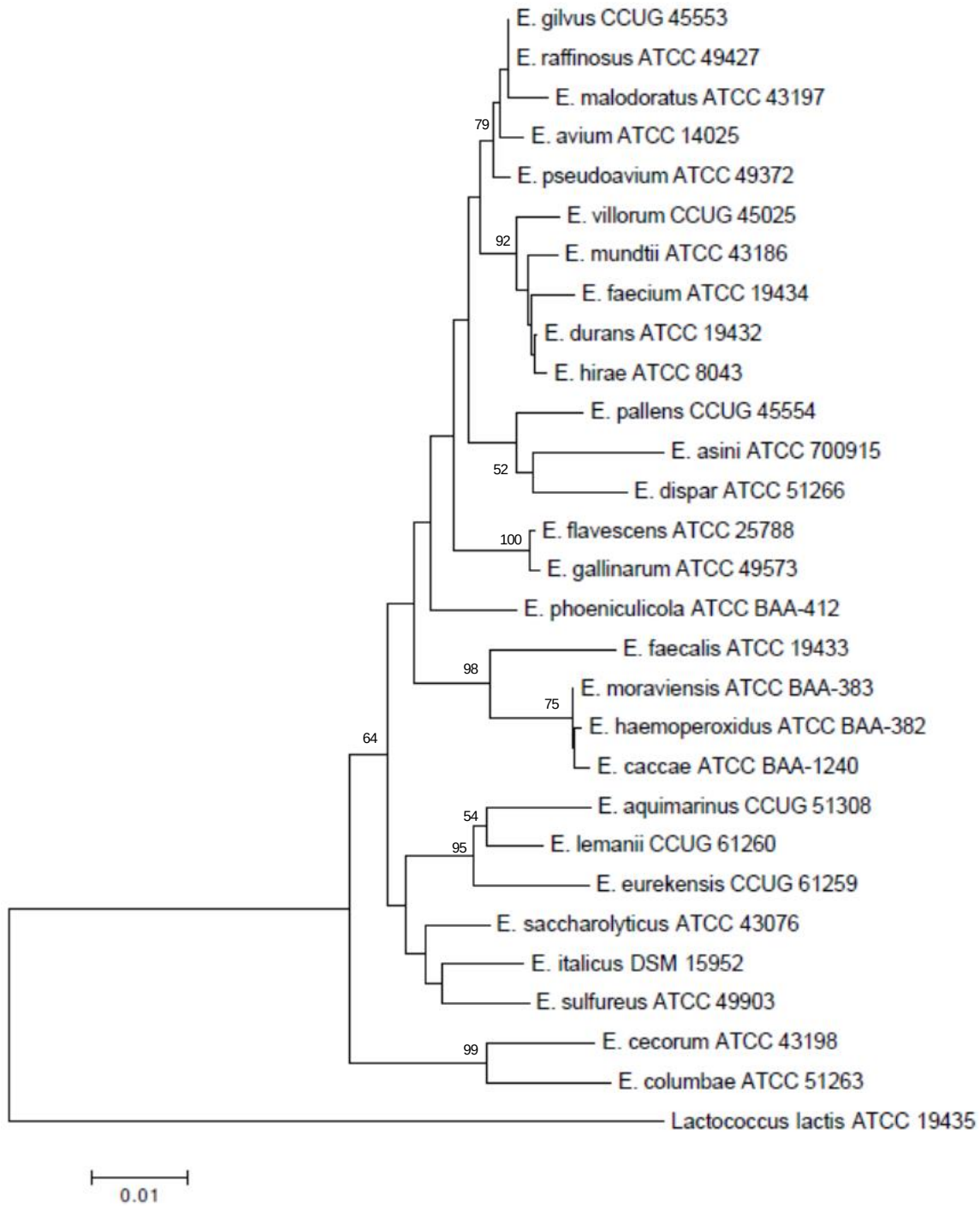


Figure 2.3 Phylogenetic tree for the genus *Enterococcus*
The tree was constructed using the 16S rRNA sequences for type strains of 28 *Enterococcus* species for which *cpn*60 UT sequence is also available. The tree was constructed using the neighbor-joining method is based on an alignment length of 1273 nucleotides. *Lactococcus lactis* is used as an outgroup. Bootstrap values were generated over 100 iterations. Bootstrap values >50 have been included in the figure.

To explore the relationship between *cpn*60 sequence based relationships and phenotypic relationships, *cpn*60 UT sequences were translated to amino acid sequences for comparison and phylogenetic analysis. The pairwise similarities between *cpn*60 UT amino acid sequences were 86-100% (Figure 2.4). The overall topologies of the peptide and nucleotide trees were similar, although three pairs of species (*E. durans* and *E. faecium*; *E. avium* and *E. malodoratus; E. cecorum* and *E. columbae*) had identical *cpn*60 UT amino acid sequences (Figure 2.4).

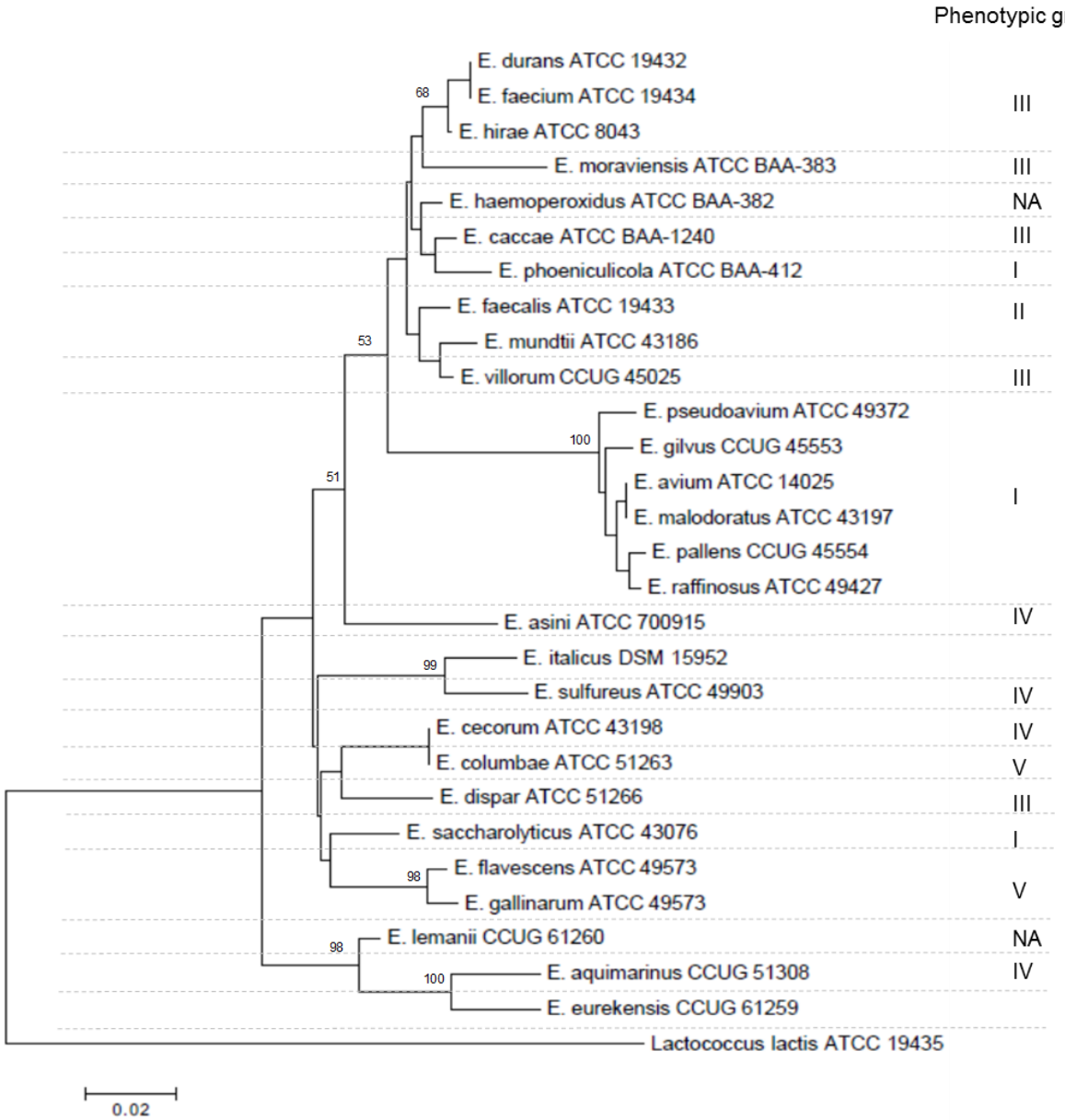


Figure 2.4 Phylogenetic tree of *cpn*60 UT peptide sequences for type strains of 28 *Enterococcus* species

The tree was constructed using the neighbor-joining method and is based on an alignment length of 184 amino acids. *Lactococcus lactis* is used as an outgroup. Bootstrap values were generated over 100 iterations. Bootstrap values >50 have been included in the figure. Roman numerals indicate phenotypic groups based on Facklam et al. (2002) NA = not available in publically available data.

Facklam et al. (2002) classified *Enterococcus* species in to five groups based on acid fermentation in mannitol and sorbose broth and hydrolysis of arginine and recently, this classification system was expanded to six groups and includes species discovered since 2002 (Lebreton et al. 2014). It is not surprising to see identical amino acid sequences for *E. avium* and *E. malodoratus* since they are both Group I members that are reported to have similar fermentation capabilities in addition to being non-motile (Figure 2.4). Similarly, *E. durans* and *E. faecium* that belong to Group III enterococci and differ only in the ability of *E. faecium* to grow in broth containing 1% arabinose have identical *cpn*60 UT sequences. The type strains for *E. cecorum* and *E. columbae* were isolated from birds (Devriese et al. 1990; Williams et al. 1989) and differ only in fermentation of arabinose and mannitol among the other sugars tested to classify these species into phenotypic groups for the genus *Enterococcus*. Further, *E. cecorum* and *E. columbae*, cannot grow or grow poorly in 6.5% NaCl broth, and cannot grow at 10°C (Devriese et al. 1993).

2.4.3 Subspecies resolution in *E. faecium* and *E. faecalis*

Our previous work demonstrated that the resolving power of the *cpn*60 UT sequence permitted the demarcation of sub-species groups within *E. faecalis* and *E. hirae* isolated from pig feces (Vermette et al. 2010) and that these *cpn*60 defined groups corresponded to ecotypes: discrete clusters within bacterial species that can be discriminated on the basis of phenotypic, ecological and DNA sequence characteristics (Cohan 2002b). *cpn*60 UT sequences have been shown to resolve phenotypically distinct sub-groups in *Gardnerella vaginalis* (Paramel Jayaprakash et al. 2012), and to predict

whole genome sequence relatedness in the genus *Thermoanaerobacter* (Verbeke et al. 2011). Based on these results, we took advantage of published collections of whole genome sequences of *E. faecalis* and *E. faecium* to determine if subspecies resolution was possible within these clinically important species.

*cpn*60 UT sequences were extracted from the whole genome sequences of 157 *E. faecium* strains (Table 2.2) originally isolated from various sources including human blood and feces, feces of pigs and chickens, and water and food products in multiple countries over a period of 54 years (1956-2010). Phylogenetic analysis was performed as explained in the methods section. Two distinct sub-groups were apparent in the *cpn*60 UT phylogeny (Figure 2.5). The differences in the alignment of *cpn*60 UT sequences between the two groups were randomly distributed and "hotspots" or variable regions that could distinguish the two clades were not evident. Pairwise nucleotide sequence identities within each sub-group were 99.1%-100% and between sub-groups were 95.6%-96.1%.

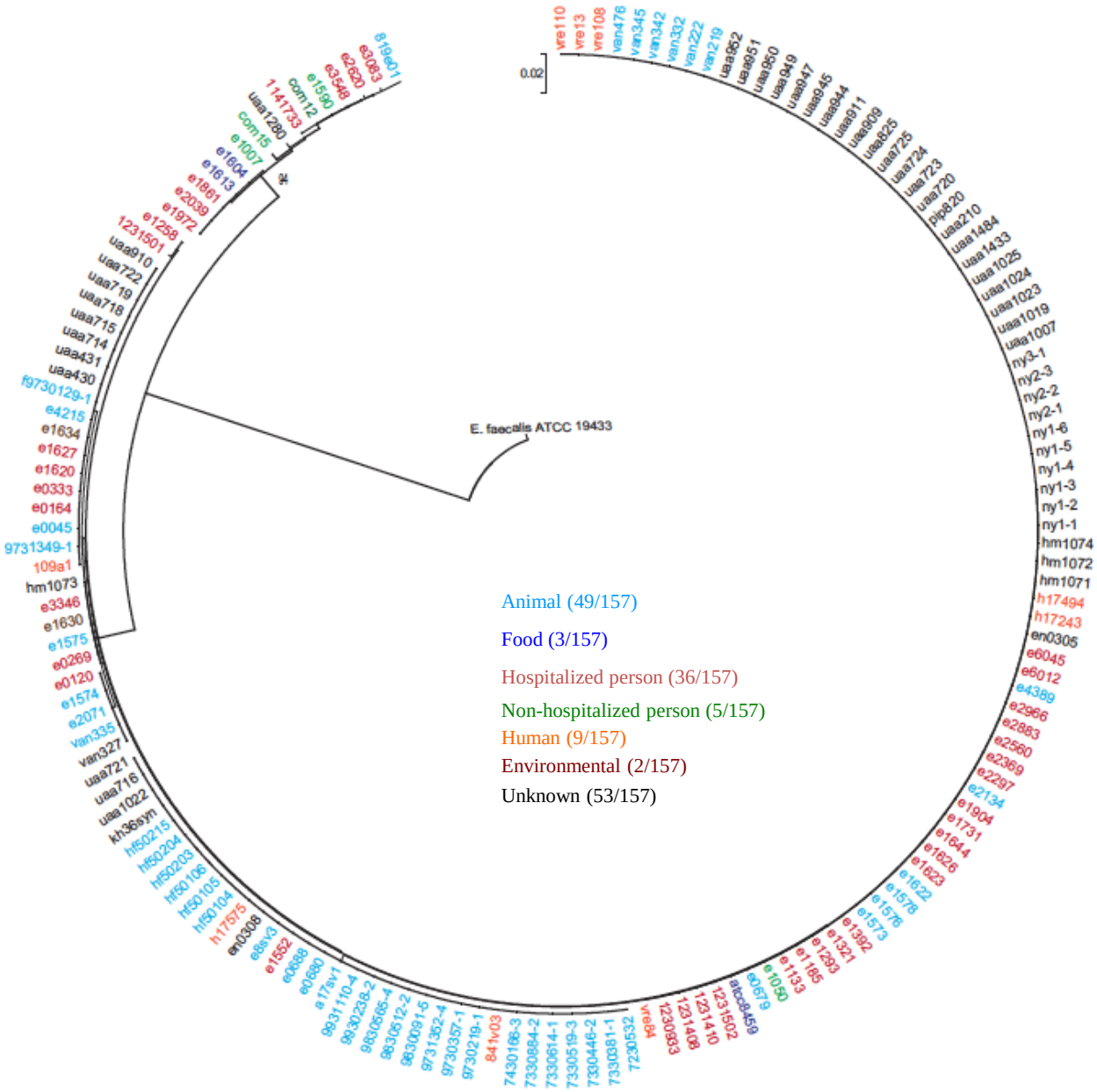

0.02
E. faecalis ATCC 19433
Animal (49/157)
Food (3/157)
Hospitalized person (36/157)
Non-hospitalized person (5/157)
Human (9/157)
Environmental (2/157)
Unknown (53/157)

Figure 2.5 Phylogenetic tree of 157 *Enterococcus faecium* isolates based on cpn60 UT nucleotide sequences

The tree was constructed using the neighbor-joining method and is based on an alignment length of 552 nucleotides. *Enterococcus faecalis* ATCC 19433 is used as an outgroup. Bootstrap values were generated over 100 iterations. Bootstrap values >50 have been included in the figure. Isolation sources are indicated by color according to the legend.

The distinctness of the sub-groups could also be seen in the non-overlapping, bimodal distribution of pairwise distances within and between *E. faecium* sub-groups (Figure 2.6). Similar values for inter and intra sub-group distances were reported in previously published analyses of 100 genes from 21 *E. faecium* genome sequences (Galloway-Pena et al. 2012) and 847 single-copy core genes from 8 *E. faecium* isolates (Palmer et al. 2012).

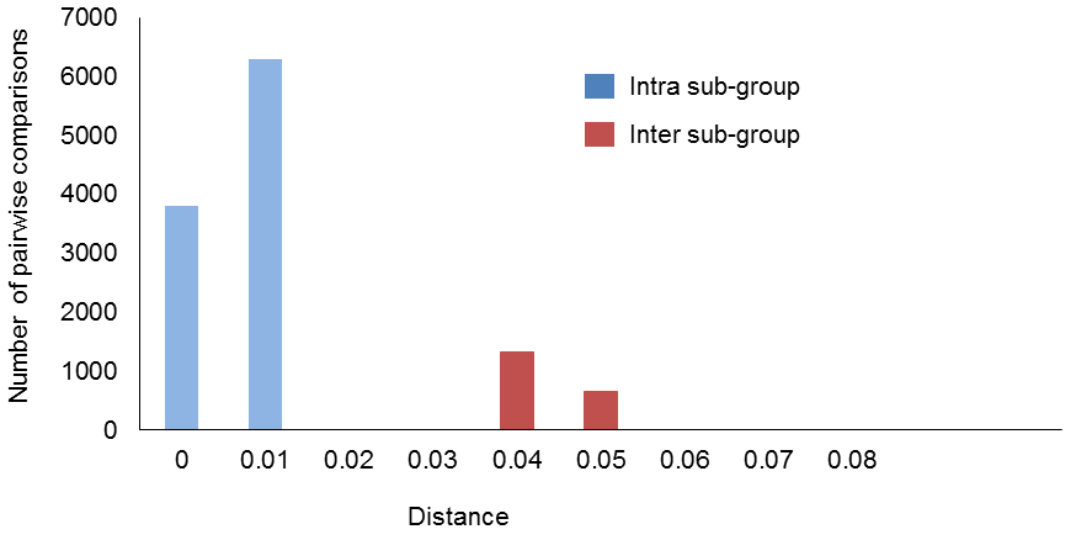
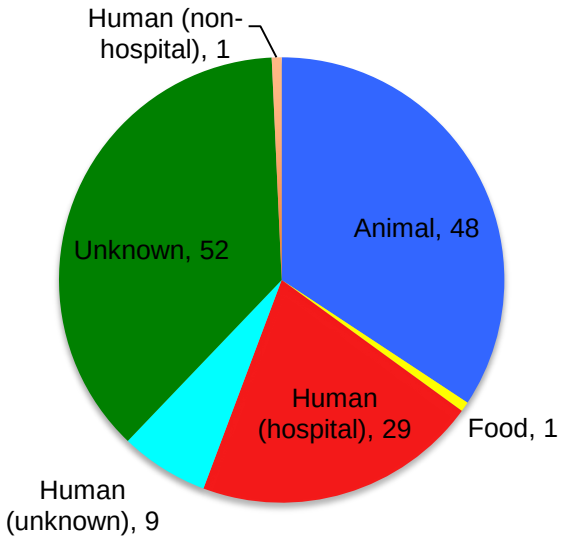


Figure 2.6 Distribution of pairwise *cpn*60 UT sequence distances for the 157 *E. faecium* isolates included in the phylogenetic analysis

The results represent a total of 12,404 comparisons. Distances for both inter-subgroup comparisons (red bars) and intra-subgroup comparisons (blue bars) are indicated.

In the two most recent comparative genomic studies by Palmer et al. (2012) and Galloway-Pena et al. (2012) it was shown that *E. faecium* isolates clustered into two distinct sub-groups: hospital associated (HA or clade A) and commensal (CA or clade B). Distinction into these sub-groups or "clades" has also been seen in other studies based on comparative genomics array analysis, amplified fragment length polymorphism (AFLP) and whole genome sequencing (Leavis et al. 2007; van Schaik et al. 2010). However, in the current study which combines sequences from both the above mentioned studies (Galloway-Pena et al. 2012; Palmer et al. 2012) neither of the two *cpn*60 UT based sub-groups of *E. faecium* were exclusive to hospital-associated strains, strains from food, animals, water or non-hospital associated strains (Figure 2.7a and 2.7b). Sub-group 1 includes a mix of isolates from animal (48/157, 30%), human hospital associated strains (29/157, 18%), non-hospitalized human (1/157, 1%) and other human associated strains (9/157, 6%), food associated strains (1/157, 1%). Only one isolate each from cheese and a non-hospitalized human were affiliated with sub-group 1. The *cpn*60 UT defined sub-group 2 comprises seven, four, two and one strain obtained from hospital sources, non-hospital sources food products and animal feces, respectively, suggesting that sub-group 2 is not limited to non-hospital associated isolates as previously reported (Galloway-Pena et al. 2012; Palmer et al. 2012). The isolation sources of 53/157 (33%) isolates could not be found in publically available literature.

A



B

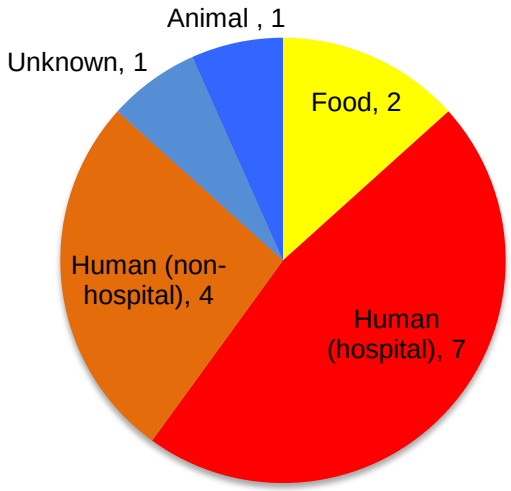


Figure 2.7 Source of isolation of *E. faecium* isolates
A. Isolation sources of 142 *E. faecium* strains that comprise *cpn*60 UT sequence defined sub-group 1. B. Isolation sources of 15 *E. faecium* strains that comprise *cpn*60 UT sequence defined sub-group 2.

Our observations are consistent with results from a study by Lebreton et al. (2013) where single nucleotide polymorphisms in 1344 shared single copy core genes from 51 *E. faecium* isolates were analyzed. Similar to our results, two *E. faecium* clades were observed (designated A and B). One of the clades included strains from animals (pets and farm animals) and hospital-associated strains, with one strain from a non-hospitalized person and four isolates from food and environmental sources. Six of 15 isolates in clade B were hospital associated. Five and three of 15 isolates in clade B were associated with non-hospitalized persons and food resources, respectively. Similar observations were also made by Willems et al. (2012), where the population structure of 1,720 isolates belonging to 519 MLST sequence types (491 for *E. faecium* and 28 for *E. faecalis*) was studied by using Bayesian-based population genetic modeling implemented in Bayesian Analysis of Population Structure (BAPS) software. In this study, *E. faecium* isolates grouped into 6 BAPS groups and 13 BAPS sub-groups. Isolates from human clinical cases, animal hosts, non-hospital human sources, food products and other environmental strains were found in all the BAPS groups and a clear distinction into hospital associated and commensal isolates was not seen.

Unlike *E. faecium*, the pairwise nucleotide percent similarities for *cpn*60 UT sequences for 182 *E. faecalis* isolates included in the study were all 98.35%-100% and no phylogenetically distinct sub-species groups were observed. The sources of isolation of these isolates are shown in Figure 2.8. Lack of phylogenetically significant sub-group resolution in *E. faecalis* has been seen in previous studies using genome sequences and MLST profiling. For example, 51 *E. faecalis* isolates obtained from different sources (hospitalized patients, non-hospitalized persons and animals) were shown to share

sequence types and group together in common clonal complexes (Palmer et al. 2012; Ruiz-Garbajosa et al. 2006; Willems et al. 2012). This could be due to the fact that genotypic variation in *E. faecalis* isolates is mainly accounted for by mobile genetic elements (Solheim et al. 2009) that are not detected by tools currently available to study evolutionary relationships between *E. faecalis* isolates.

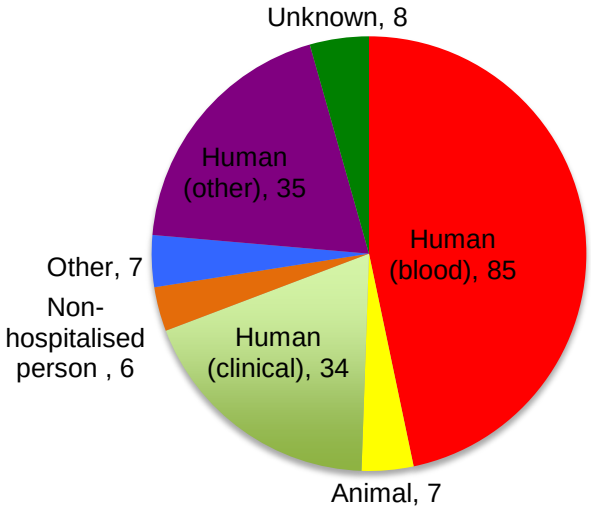


Figure 2.8 Source of isolation for 182 *E. faecalis* strains used in the study

It is interesting to note that the *E. faecium* sub-groupings observed by phylogenetic analysis of *cpn*60 UT sequences are consistent with results obtained from whole genome sequence analysis. The *cpn*60 UT can be found in virtually all bacteria and can be amplified with a set of universal amplification primers. The UT sequences are short enough that they can be sequenced in a single reaction using di-deoxy sequencing methods, but provide better resolution than full-length 16S rRNA sequences which are >1kb in length. These features, combined with previous evidence of whole genome similarity prediction (Paramel Jayaprakash et al. 2012; Verbeke et al. 2011) make *cpn*60 UT based sequencing a rapid, simple and cost-effective complement to whole genome sequencing for identification of bacterial species and subspecies and investigating bacterial population structure. For population structure investigation, *cpn*60 UT sequencing could provide a valuable screening tool for initial investigations of isolates prior to undertaking the work and expense of whole genome sequencing.

Taken together, the results of this study show that *cpn*60 UT sequences can be used to differentiate and identify *Enterococcus* species, and resolve sub-species groups in *E. faecium.* With increased exploration of microbial environments, identification of new enterococcal species and involvement of enterococci in hospital-associated infections, *cpn*60 UT should be used as a first step for identification and screening of enterococci in large-scale studies where genome sequencing and/or subspecies resolution are a consideration.

CHAPTER 3 - Characterization of genomic diversity in *Enterococcus* ecotypes using GTG rep PCR

Third chapter transition

In our previous studies, it was seen that *cpn*60 UT sequences could resolve sub-species groups within *E. faecalis*, and *E. hirae* isolates derived from pig feces. Results described in the previous chapter indicate that this resolution is also observable beyond the pig *Enterococcus* community. Whole genome sequencing studies of human clinical isolates of *E. faecium* and *E. faecalis* suggest significant variation in genome content among isolates, but the extent to which *cpn*60 based resolution corresponds to genomic diversity in enterococci is not known. The following chapter describes the results of analysis of a collection of 108 *E. hirae* and 82 *E. faecalis* porcine isolates using a genome fingerprinting method, GTG rep PCR, with the goal of determining if *cpn*60 based subgroups are consistent with clustering based on whole genome structure. Demonstration of genome-based clustering of these isolates that is consistent with previously observed *cpn*60-based patterns would strengthen the hypothesis that these groups represent ecotypes: populations of related strains that occupy distinct ecological niches.

3.1 Abstract

Enterococci form an important part of intestinal bacterial community in animals and include both pathogenic and non-pathogenic species. Previous work demonstrated a shift in the fecal *Enterococcus* community composition in healthy pigs with change in diet over time, suggesting that selection (due to different diets and age of the pig) plays an important role in community development and succession at the species and subspecies level. It was also shown that *cpn*60 UT sequences could resolve ecotypes (genotypically and phenotypically distinct strains within a species) within *E. faecalis* and *E. hirae*. In this study, we investigated genomic diversity among pig fecal *Enterococcus* ecotypes using (GTG) repetitive extragenic palindrome polymerase chain reaction (GTG rep PCR). Genomic DNA extracted from cultured *Enterococcus* isolates was used to generate GTG rep PCR profiles for comparison between and within *cpn*60 defined sub-groups and type strains. *E. faecalis* and *E. hirae* isolates could be distinguished based on their GTG rep PCR based genomic fingerprints. The minimum percent similarity between the fingerprints of *E. faecalis* and *E. hirae* isolates was 43.7% and 59.8%, respectively. GTG rep PCR based clusters for *E. faecalis* and *E. hirae* isolates were generally consistent with the *cpn*60 defined *Enterococcus* ecotypes observed in our previous study suggesting that *cpn*60 UT is a predictor of genomic relationships. Our results show that GTG rep PCR is a useful technique to detect genome level differences between and within *Enterococcus* ecotypes that may be due to genome re-arrangements and/or genome content differences.

3.2 Introduction

Enterococci are ubiquitous Gram positive organisms commonly isolated from intestines of animals (Devriese et al. 1992a; Devriese et al. 1992b; Devriese et al. 1994; Devriese et al. 1992c), food products (Suzzi et al. 2000; Yousif et al. 2005), insects (Martin and Mundt 1972) and the environment (Mundt 1961; Svec and Sedlacek 1999). Often considered a serious issue in human medicine because of their antibiotic resistance and frequent association with hospital acquired infections, this group represents an interesting paradox as it includes both pathogenic and non-pathogenic species. In spite of numerous studies of genomic diversity among commensal and pathogenic isolates, and environmental *Enterococcus* spp., little investigation of subspecies diversity of enterococci in natural microbial communities has been done. Their relative ease of culture and isolation and abundance in intestinal communities make enterococci an ideal population to target in studies of microbial diversity and succession.

A number of *Enterococcus* species have been identified in animal intestines and *E. faecalis, E. faecium, E. hirae, E. mundtii* and *E. avium* have been reported among the most commonly occurring species in the pig intestine (Lebreton et al. 2014). In our previous study, we established a culture collection of 196 *Enterococcus* isolates including *E. faecalis, E. faecium, E. avium, E. mundtii* and *E. hirae* from fecal samples of healthy pigs (n=14) at 3, 9, and 15 weeks of age (Vermette et al. 2010). Piglets at 3 weeks were suckling. Diet composition at 9 weeks consisted of 55 % wheat, 26 % soymeal, and 15 % barley. At 15 weeks, the diet consisted of 30 % corn, 23 % barley, 14 % soymeal, 8 % canola meal, and 23 % wheat. We documented changes in the fecal *Enterococcus* community of growing pigs, with an overall decrease in population size, and increase in

species richness between 3 and 15 weeks of age . Further, isolates from this environment could be grouped into phylogenetically distinct subspecies groups using *cpn*60 UT sequencing. Also, these subspecies groups had different sole carbon source utilization profiles, suggesting that they represent ecotypes: groups of closely related strains occupying the same ecological niche (Cohan 2002b). These subspecies level groupings were not detected with 16S rRNA sequences. Succession was observed at the ecotype level, reflecting change in diet and increasing age of the animal. This observation was significant in that it shows that succession is occurring in this community at the subspecies level, which would not be detected using common culture-independent methods based on sequencing the 16S rRNA gene. Similarly, in the previous chapter we showed that the *cpn*60 UT sequence could resolve species in the genus *Enterococcus* and subgroups in *E. faecium*. A better understanding of the potential biological significance of subspecies level diversity revealed by *cpn*60 UT sequencing depends upon characterization of genome level characteristics of the *cpn*60 defined ecotypes.

A number of genome fingerprinting techniques have been developed for use in identification of species (De Vuyst et al. 2008; Masco et al. 2003), epidemiological investigations (Ostojic 2008; Watabe et al. 2008) and microbial source tracking (Valdezate et al. 2007). Among the techniques available, pulse field gel electrophoresis (PFGE) is considered a gold standard for genome fingerprinting owing to its high discriminatory power (Bosch et al. 2013; Karden-Lilja et al. 2013; Tenover et al. 1997). PFGE has been used to identify and classify genetic variants for a number of bacteria including pathogenic strains of *E. faecalis* and *E. faecium* (Gordillo et al. 1993). However, PFGE is time consuming and requires expensive and specialized equipment.

Alternatives to PFGE include PCR based methods that target repetitive elements in bacterial genomes such variable number tandem repeat (VNTR) analysis, multi-locus variable number tandem repeat analysis (MLVA) and repetitive element based PCR (Struelens 1998). These methods are based on the principle that specific primers for repeated sequences in the genome will amplify strain specific DNA fragments that can then be resolved by electrophoresis to obtain a genome fingerprint.

GTG rep PCR, a type of genome fingerprinting technique, utilizes a single $(GTG)_5$ oligonucleotide primer that enables simultaneous amplification of multiple, different sized DNA fragments between GTG repetitive sequences in the genome. These interspersed repetitive sequences occur in relatively high copy number and are conserved in diverse genera of bacteria. This method has been used for investigating relationships between strains of several bacterial groups including lactobacilli, staphylococci, mycobacteria, and streptomycetes (Cangelosi et al. 2004; Gevers et al. 2001; Wieser and Busse 2000). GTG rep PCR has been used for studying genomic diversity between *Enterococcus* isolates from clinical and environmental sources (Svec et al. 2005c) suggesting that it is an appropriate technique for investigating genomic relationships between *cpn*60 defined ecotypes of *Enterococcus* isolated from pig feces.

The objective of the current study was to determine if subspecies groups defined by *cpn*60 UT based phylogenetic analysis can also be resolved based on whole genome fingerprints generated by GTG rep PCR. The results of the study may also provide a basis for selection of isolates for further gene level characterization by whole genome sequencing.

3.3 Material and methods

3.3.1 Bacterial isolates

Enterococcus isolates were obtained from fecal samples of healthy pigs as described previously (Vermette et al. 2010) and named according to pig number-age-isolate number. A subset of 190 *E. faecalis* (n=82) and *E. hirae* (n=108) isolates from the original 694 isolates was selected to represent all unique *cpn*60 UT sequences (Table 3.1). Type strains of *E. faecalis* (ATCC 19433) and *E. hirae* (ATCC 8043) were obtained from American Type Culture Collection (Manassas, VA). All isolates were cultivated in tryptone soy broth (TSB), incubated aerobically at 35 ºC.

Table 3.1 Number of isolates representing *cpn*60 defined ecotypes isolated from feces of healthy pigs at 3, 9 and 15 weeks of age.

***cpn*60 defined ecotype**	**Number of isolates (number of pigs represented)**			**Total**
	3 week	9 week	15 week	
E. hirae-1	0 (0)	62 (7)	1 (1)	63
E. hirae-2	0 (0)	0 (0)	39 (4)	39
E. hirae-3	2 (3)	3 (2)	0 (0)	5
E. hirae-4	0 (0)	0 (0)	1 (1)	1
E. faecalis-1	11 (2)	0 (0)	0 (0)	11
E. faecalis-2	43 (4)	10 (3)	16 (6)	69
E. faecalis-3	1 (1)	1 (1)	0 (0)	2
Total	57	76	57	190

3.3.2 Genomic DNA extraction

Genomic DNA was extracted separately from duplicate 2 mL broth cultures of type strains and pig isolates using a modified salting out procedure as previously described (Martin-Platero et al. 2007). Extracts were stored at -20 ºC in TE buffer (1 mM EDTA, 10 mM Tris-HCl, pH 7.4). The quality of DNA extracts was assessed by spectrophotometric measurements at 260 and 280 nm. A 260/280 ratio of 1.8-2.0 was considered good. DNA extracts were also evaluated by gel electrophoresis in 1% agarose in 0.5× TBE (44.5 mM Tris, 44.5 mM boric acid, 1mM EDTA, pH 8.3). A high molecular weight band with minimal smearing was considered good.

3.3.3 GTG rep PCR

GTG rep-PCR was performed as described previously (Versalovic et al. 1991). Each 50 µl reaction mixture contained 1 × PCR buffer (100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris HCl pH 8.75, 1% (v/v) Triton X-100, 1 mg/ml BSA) (UBI Life Sciences, Saskatoon, SK), 2 mM $MgSO_4$, 50 pmol of primer $(GTG)_5$ (5'-GTG GTG GTG GTG GTG-3'), 10% (v/v) dimethyl sulfoxide, 0.2 mM dNTPs (Invitrogen, Mississauga, ON), and 2.5 U HP Taq polymerase (UBI Life Sciences, Saskatoon, SK). PCR amplifications were performed in an Eppendorf Mastercycler EP gradient thermocycler with an initial denaturation step (95 ºC for 3 minutes), followed by 35 cycles of denaturation (95 ºC, 30 seconds), annealing (40 ºC, 1 minute) and extension (65 ºC, 8 minutes) and a single final extension step (65 ºC, 10 minutes). Products were resolved by gel electrophoresis in a 1% (w/v) agarose gel for 5 hours at 100V in 0.5× TBE (45 mM Tris base, 45 mM boric acid, 1 mM EDTA pH 8.0). Gels were photographed using an

Alpha Imager gel documentation system and saved as .TIFF files for analysis with GelCompar II V6.1 software (Applied Maths Inc., Austin, TX). Banding patterns were clustered using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering of dice co-efficient with optimization and position tolerance settings of 1% and 1.5% each.

3.4 Results and Discussion

GTG rep PCR and other related rep PCR techniques have been used for *Enterococcus* species identification (Pangallo et al. 2008), and studying genomic diversity of *Enterococcus* isolates from clinical and environmental sources (Svec et al. 2005c) but characterization of genomic diversity in commensal intestinal *Enterococcus* isolates has not been reported. PFGE is the gold standard for studying sub-species resolution in several bacterial species including *Enterococcus* (Karden-Lilja et al. 2013; Weng et al. 2013). However, more simple and rapid fingerprinting methods may be advantageous in some cases, especially where comparison of large numbers of isolates is needed. GTG rep PCR has been used for identification of *Enterococcus* species and has been shown to be a more rapid and informative method for identification of *Enterococcus* species (Svec et al. 2005c) compared to other PCR based fingerprinting methods (ERIC, REP and BOX PCR). In this study, we applied GTG rep PCR to determine the genomic diversity between *cpn*60 UT defined *Enterococcus* ecotypes.

Although discriminatory power is an important parameter for evaluating fingerprinting methods, repeatability of repetitive element based PCR (rep PCR) methods is equally important. Lack of reproducibility of rep PCR methods has been shown to be a

concern and has been attributed primarily to qualitative differences in banding patterns of the genomic fingerprints (Johnson and O'Bryan 2000). Several factors such as differences in thermocycler, PCR reagents, DNA extraction methods (conventional DNA extraction methods vs. commercial kits) and electrophoresis conditions can account for variability in genomic fingerprints obtained from rep PCR based methods (Johnson and O'Bryan 2000). Use of intact chromosomal DNA as a template for these techniques is also critical, since highly fragmented genomic DNA template would yield fewer bands and correspondingly less informative fingerprints. Studies evaluating the repeatability of rep PCR based fingerprinting methods generally have not been supported with specific data (usually "data not shown") or rely only on a limited number of isolates (De Vuyst et al. 2008; Mohapatra et al. 2007), reducing the strength of the conclusions of these studies.

In the current study, the repeatability of GTG rep PCR was assessed by extracting genomic DNA from duplicate cultures of all the isolates and performing GTG rep PCR on all extracts. Fingerprints were obtained by gel electrophoresis of the PCR product and analyzed using GelCompar IIGCII software as described in the methods section. Examples of duplicate fingerprints for two *E. faecalis-2* isolates (81-9-A1 and 81-3-B1) and four *E. hirae-2* isolates (88-15-H6, 81-15-F4, 88-15-E9, 81-15-F1, 88-15-D6 and 88-15-D3) are shown in Figure 3.1. It can be seen that the duplicate fingerprints for all of these *E. faecalis* and *E. hirae* isolates have identical banding patterns. The relationship between duplicate fingerprints for all 190 study isolates and the type strains was determined by using UPGMA clustering of dice co-efficient with optimization and position tolerance settings of 1% and 1.5% each. Dice co-efficient was chosen over the

Pearson co-efficient as it determines the relationship based on presence or absence of bands in a fingerprint rather than the band intensity in the fingerprints.

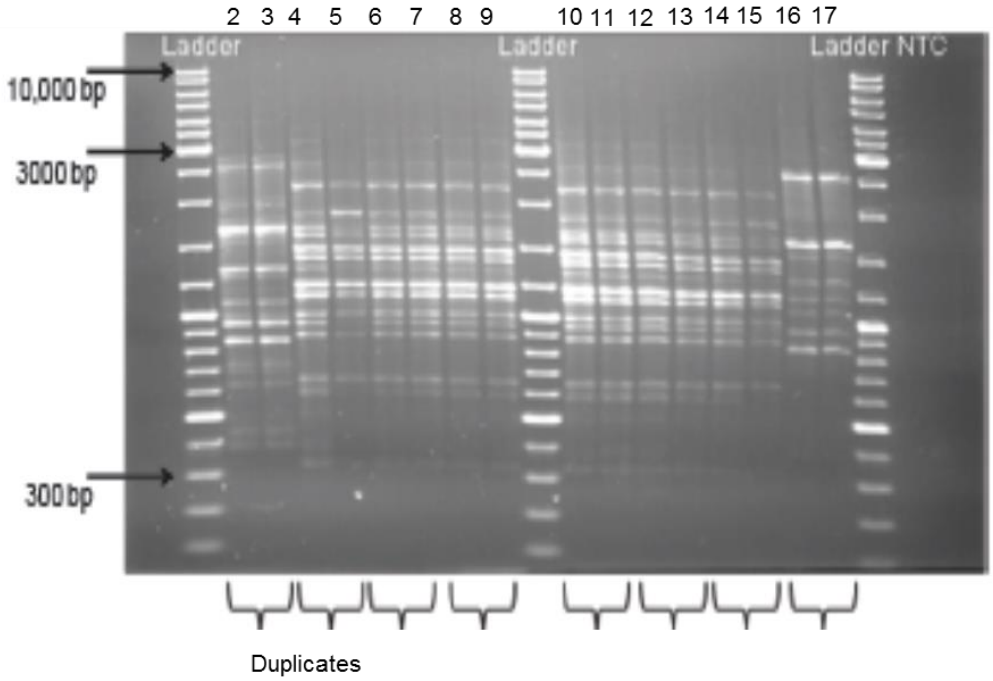


Figure 3.1 GTG fingerprints for *Enterococcus* isolates
81-9-A1 (*E. faecalis*-2, lanes 2 and 3), 88-15-H6 (*E. hirae*-2, lanes 4 and 5), 81-15-F4 (*E. hirae*-2, lanes 6 and 7), 88-15-E9 (*E. hirae*-2, lanes 8 and 9), 81-15-F1 (*E. hirae*-2, lanes 10 and 11), 88-15-D6 (*E. hirae*-2, lanes 12 and 13), 88-15-D3 (*E. hirae*-2, lanes 14 and 15), 81-3-B01 (*E. faecalis*-2, lanes 16 and 17); NTC: No Template Control

The average percent similarity between duplicate fingerprints for 170/190 (90%) isolates was 92% (range 85-100%), which is consistent with previous studies that have evaluated repeatability of GTG rep PCR in smaller collections of isolates (De Vuyst et al. 2008; Klima et al. 2010). In our study, poor duplicates (10%, 20/190) were attributable to one replicate having faint bands or no bands. Similar observations were made by Svec et al. (2005c) and Gevers et al. (2001) who evaluated the repeatability of GTG rep PCR for identification of *Enterococcus* and *Lactobacillus* isolates, respectively, and showed that differences between replicate rep PCR fingerprints were largely accounted for by variation in band intensities, rather than qualitative differences in banding patterns. In our study, even though the banding patterns in duplicate fingerprints were often visually identifiable to contain bands of the same size, differences in band intensities resulted in misidentification of bands when identified using the GelCompar II software. Although visual inspection of fingerprints and manual annotation of bands improves repeatability of results, the time saving attributes associated with PCR based techniques in comparison to PFGE are reduced and the influence of observer bias increases. Fingerprints of the 170 isolates with ≥85% similar duplicate fingerprints (*E. faecalis*: n=69, *E. hirae*: n=101) were used in further analysis.

A dendrogram representing the relationships between all the *E. faecalis* (n=69) and *E. hirae* (n=101) isolates and the respective type strains is shown in Figure 3.2. The minimum percent similarity between fingerprints of all the isolates was 18.7%. GTG rep PCR fingerprints of *E. faecalis* (63/69) cluster together with the *E. faecalis* type strain with minimum percent identity of 44.1% (Figure 3.2). Similarly, (89/101) *E. hirae* isolates clustered with the *E. hirae* type strain with minimum percent identity of 45%

(Figure 3.2). However, 6/69 (8%) *E. faecalis* and 12/101 *E. hirae* (11%) formed distinct clusters. These results are generally consistent with previous reports where GTG rep PCR was used for identification of 44 previously unknown *Enterococcus* isolates obtained from water (Svec et al. 2005c) where isolates clustered according to species.

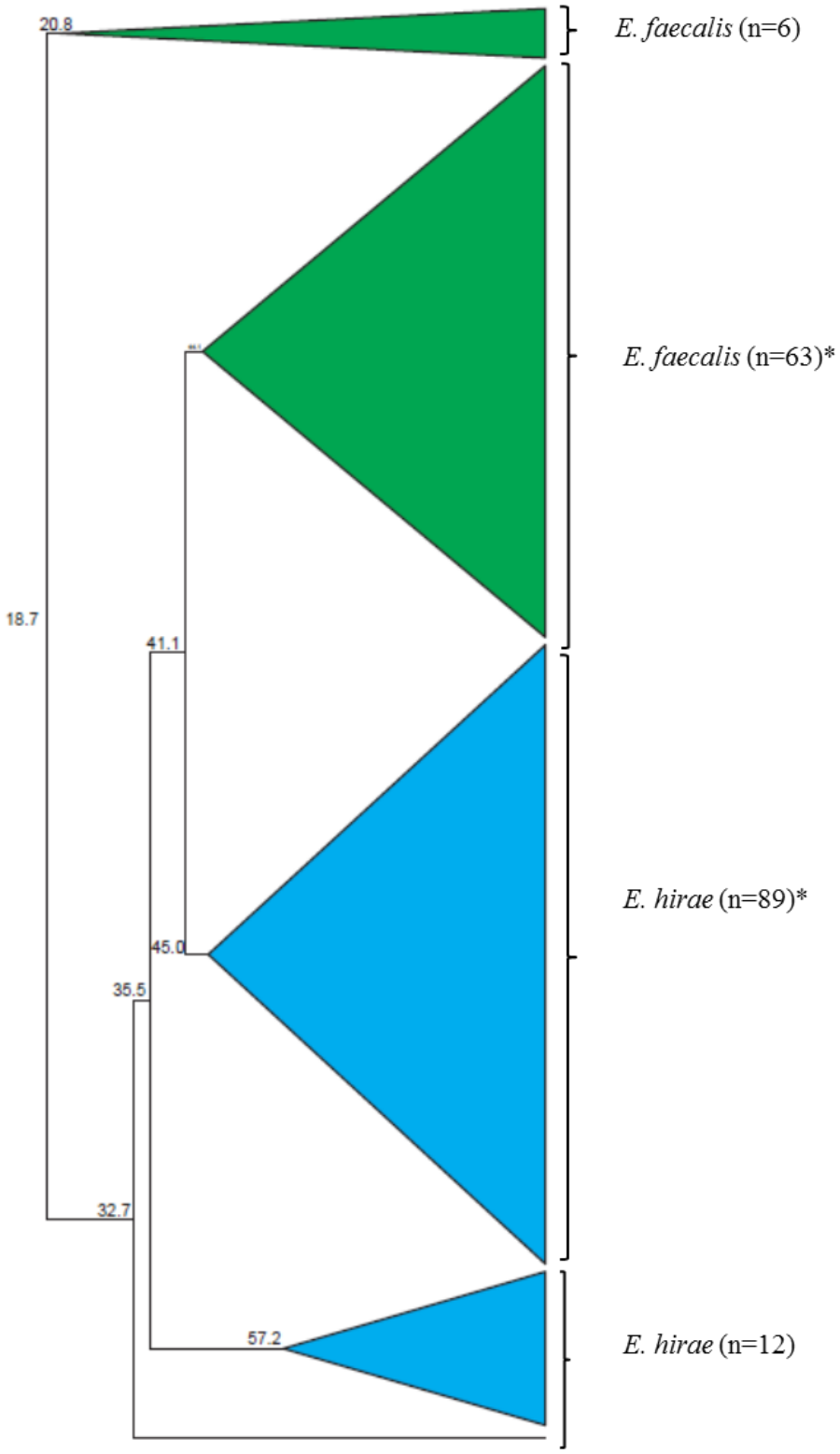

20.8
18.7
41.1
45.0
35.5
32.7
57.2
E. faecalis (n=6)
E. faecalis (n=63)*
E. hirae (n=89)*
E. hirae (n=12)

Figure 3.2 UPGMA clustering of GTG fingerprints of all *Enterococcus* isolates and type strains.
Fingerprints have been collapsed at the last common node for each cluster. Numbers on the branches indicate percent similarity. The clusters containing the type strains for *E. faecalis* and *E. hirae* are indicated by *.

To examine the GTG rep PCR fingerprints of ecotypes, fingerprints of *E. hirae* and *E. faecalis* fecal isolates were evaluated separately. Fingerprints of isolates within ecotypes of *E. hirae* (*E. hirae-1* and *E. hirae-2*) were more similar to each other than between the two groups. Clustering of genome fingerprints of the all the *E. hirae* ecotypes: *E. hirae-1* (n=59) and *E. hirae-2* (n=37), *E. hirae-3* (n=4) and *E. hirae-4* (n=1) were generally consistent with *cpn*60 UT based clustering (Figure 3.3). The minimum percent similarity between the fingerprints of all the *E. hirae* isolates was 59.8%. The percent similarities for four distinct clusters comprising fingerprints from 17, 8, 4 and 5 *E. hirae-2* isolates, respectively, ranged from 73.5% to 100%. The fingerprints of isolates from *E. hirae-3* and *E. hirae-4* ecotypes did not cluster into distinct groups as observed in the *cpn*60 UT based analysis.

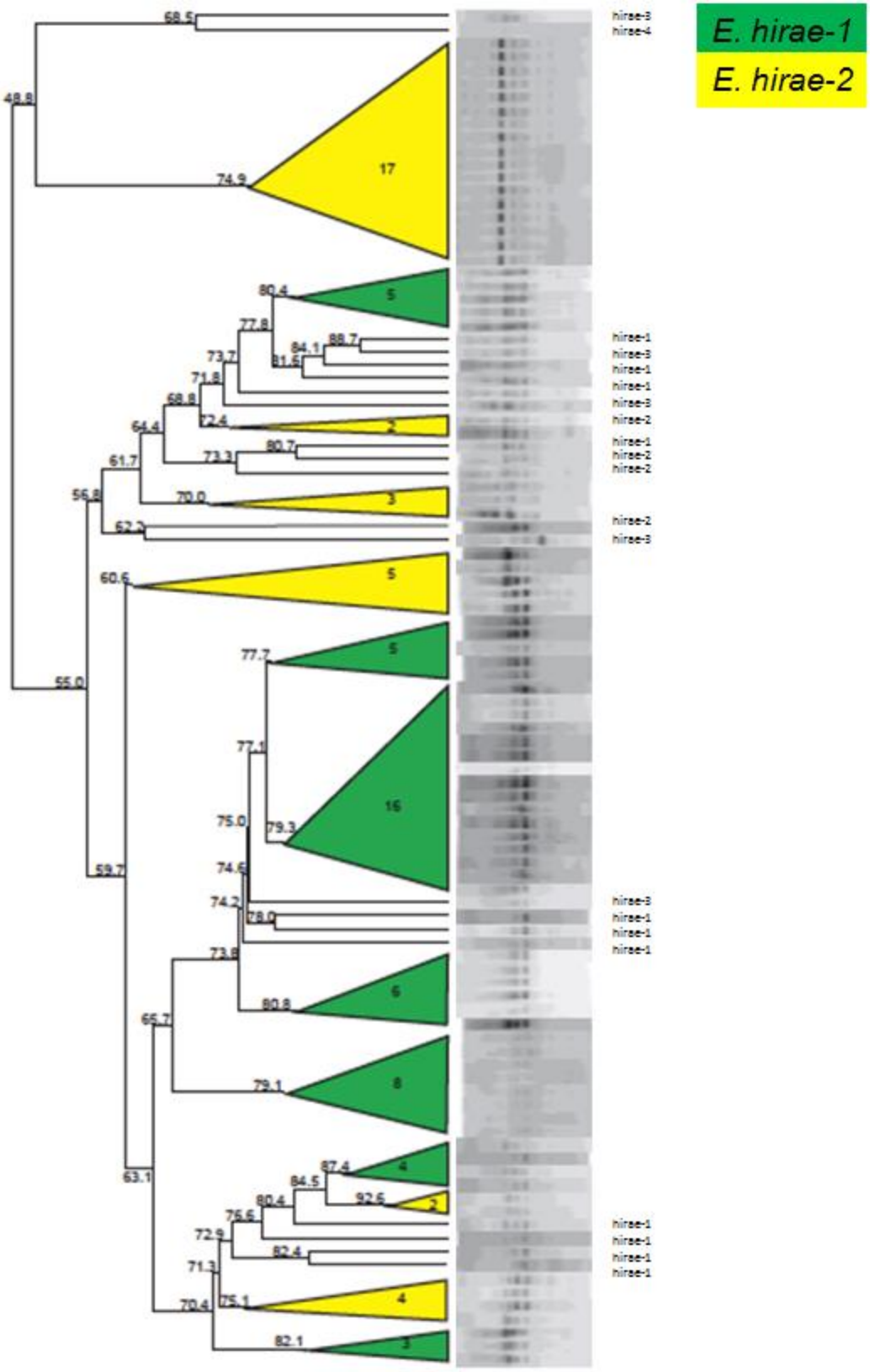

E. hirae-1
E. hirae-2
hirae-3
hirae-4
hirae-1
hirae-2
58.5
48.8
74.9
17
80.4
5
77.8
84.1
88.7
73.7
81.6
71.8
68.8
72.4
2
64.4
80.7
73.3
61.7
56.8
70.0
3
62.2
60.6
5
77.7
5
55.0
77.1
16
75.0
79.3
59.7
74.6
74.2
78.0
73.8
80.8
6
65.7
79.1
8
87.4
4
84.5
92.6
2
80.4
76.6
72.9
82.4
71.3
75.1
4
70.4
82.1
3
63.1

Figure 3.3 UPGMA clustering of GTG fingerprints of all *E. hirae* isolates.
Fingerprints have been collapsed at the last common node for each sub-group cluster. Numbers on the branches indicate percent similarity and numbers in the triangles indicate number of isolates in each cluster.

Clustering of genome fingerprints of *E. faecalis* ecotypes: *E. faecalis-1* (n=8) and *E. faecalis-2* (n=61), was generally consistent with the *cpn*60 UT defined ecotypes (Figure 3.4). The correspondence between GTG rep PCR based *E. faecalis* clusters with the *cpn*60 UT sequence based analysis may also be due to the number greater of *E. faecalis-2* (n=61/69) in comparison to *E. faecalis-1* (n=8/11) isolates in our culture collection, and the fact that *E. faecalis-1* isolates were obtained only from two pigs (pig 68, six isolates and pig 88, two isolates). The minimum percent similarity between the fingerprints of *E. faecalis* isolates was 43.7%. A distinct cluster including 7/8 *E. faecalis-1* isolates was seen. The percent similarity for fingerprints of *E. faecalis-1* isolates ranged from 78% to 100%. Fingerprints from *E. faecalis-2* isolates grouped in to two major clusters comprising 25 isolates each. A third cluster comprising fingerprints from eight *E. faecalis-2* isolates was also seen. The percent similarity for fingerprints of *E. faecalis-2* isolates ranged from 65.2% to 95.2%.

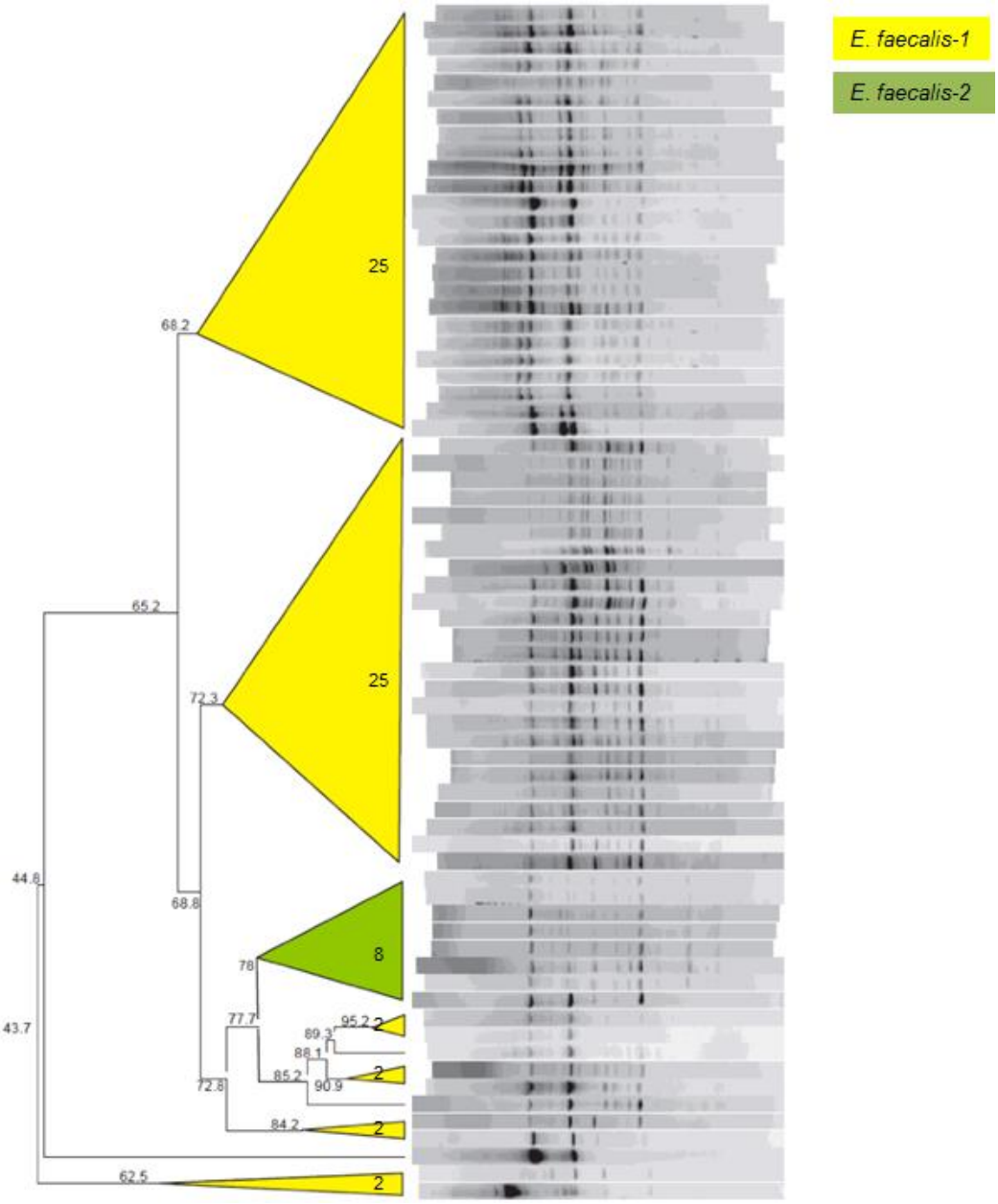


Figure 3.4 UPGMA clustering of GTG fingerprints of all *E. faecalis* isolates.
Fingerprints have been collapsed at the last common node for each sub-group cluster. Numbers on the branches indicate percent similarity and numbers in the triangles indicate number of isolates in each cluster.

A perfect correspondence between the *cpn*60 UT based grouping and GTG rep PCR based grouping was not observed as these tools detect different aspects of the genome. *cpn*60 is a conserved gene that provides valuable information about the phylogenetic relationships between bacteria, is a useful tool for species identification and resolution of sub-groups within a species, but its sequence provides no information about genome arrangement or mobile genetic elements such as plasmids and phages. On the other hand, GTG rep PCR is based on the presence or absence of a non-essential repetitive DNA sequences (GTG), providing information on genome arrangement, which is influenced by presence of mobile genetic elements particularly in case of the genus *Enterococcus* that has been shown to exhibit genome plasticity.

GTG fingerprints could distinguish isolates that were identical or virtually identical by *cpn*60 UT sequence. In our earlier study (Vermette et al. 2010), we observed that isolates within a *cpn*60 UT defined sub-group are >97% identical at the nucleotide level (over 552 nucleotides of the *cpn*60 UT region). But with the use of GTG rep PCR, another level of diversity within groups of closely related isolates can be seen. For example, the *cpn*60 UT sequence of the *E. faecalis* type strain (ATCC 19433) is 99.8% identical to an *E. faecalis* isolate from pig 71 (71-3-f08), while the GTG fingerprints of these strains are only 81% similar.

It can be concluded that genome-fingerprinting techniques such as GTG rep PCR can be considered a screening tool to study genome structure differences. It is also evident that *cpn*60 UT based grouping are generally consistent with GTG rep PCR based groupings suggesting that *cpn*60 UT is a predictor of whole genome relatedness.

In this study, we observed genome level differences in phylogenetically and phenotypically distinct *Enterococcus* ecotypes. The observed genomic diversity within ecotypes could be the result of genome arrangement differences or genome content differences, or both. Further studies to identify genome content differences between *cpn*60 UT defined ecotypes are needed to obtain insights into genomic diversity in natural microbial communities over time.

CHAPTER 4 - CRISPRs of *Enterococcus faecalis* and *E. hirae* isolates from pig feces have species-specific repeats, but share some common spacer sequences

This chapter has been published as:

Isha Katyal, Bonnie Chaban, Beata Ng and Janet E. Hill. 2013. CRISPRs of *Enterococcus faecalis* and *E. hirae* isolates from pig feces have species-specific repeats, but share some common spacer sequences. Microbial Ecology, 66:182-188.

Author Contributions

Conceived and designed the experiments: IK BC JEH. Performed the experiments: IK BC BN. Analyzed the data: IK BC BN JEH. Wrote the paper: IK JEH.

Fourth chapter transition

In our earlier studies, we have shown that genotypically and phenotypically distinct subgroups of enterococci that can be distinguished by *cpn*60 UT sequence emerge due to periodic selection in the pig fecal microbiome. Results presented in the previous chapter show that *cpn*60 defined subgroups of *E. hirae* are further distinguishable by whole genome structure, as indicated by GTG rep PCR fingerprinting, while *E. faecalis* subgroups defined by *cpn*60 UT sequences were not. Substantial variation in GTG rep PCR fingerprints was observed within clusters for either species, suggesting diversity in genome structure and/or content. Factors that contribute to genomic diversity in natural bacterial populations include lateral gene transfer between community members through a variety of mechanisms, including phage infection, uptake of DNA from the environment and conjugation. Clustered regularly interspaced short palindromic repeats (CRISPR) provide an accessible record of exposure to foreign DNA and thus reflect interactions within bacterial communities. In this study, we characterized CRISPR profiles for all the *Enterococcus* isolates in our culture collection to obtain genome level insight into interactions that occur within the community and may account for genome level diversity within species.

4.1 Abstract

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) are currently a topic of interest in microbiology due to their role as a prokaryotic immune system. Investigations of CRISPR distribution and characterization to date have focused on pathogenic bacteria, while less is known about CRISPR in commensal bacteria, where they may have a significant role in the ecology of the microbiota of humans and other animals, and act as a recorder of interactions between bacteria and viruses. A combination of PCR and sequencing was used to determine prevalence and distribution of CRISPR arrays in *Enterococcus faecalis* and *E. hirae* isolates from the feces of healthy pigs. Both type II CRISPR-Cas and Orphan CRISPR (without Cas genes) were detected in the 195 isolates examined. CRISPR-Cas was detected in 52% (46/88) and 42% (45/107) *E. faecalis* and *E. hirae* isolates, respectively. The prevalence of Orphan CRISPR arrays was higher in *E. faecalis* isolates (95%, 84/88) compared to *E. hirae* isolates (49%, 53/107). Species-specific repeat sequences were identified in Orphan CRISPR arrays, and 42 unique spacer sequences were identified. Only two spacers matched previously characterized pig virome sequences, and many were apparently derived from chromosomal sequences of enterococci. Surprisingly, 17 (40%) of the spacers were detected in both species. Shared spacer sequences are evidence of a lack of species-specificity in the agents and mechanisms responsible for integration of spacers, and the abundance of spacer sequences corresponding to bacterial chromosomal sequences reflects interspecific interactions within the intestinal microbiota.

4.2 Introduction

The intestinal microbiota comprises a diverse population of prokaryotes, eukaryotes and viruses, and thus, is a "hot-spot" for genetic recombination and exchange. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) are DNA arrays containing multiple unique sequences separated by short repeats. They were observed first in *E. coli* in 1987 (Ishino et al. 1987) although their significance was not immediately apparent. Since then, CRISPR arrays have been identified in approximately 40% of Bacteria and 90% of Archaea (Kunin et al. 2007; Sorek et al. 2008). Generally, a CRISPR array contains 4–10 repeats ranging in size from 25-45 bp, separated by sequences of similar length (spacers) containing unique genomic material that is apparently sampled from invading plasmids or bacteriophages, or other foreign DNA (Bolotin et al. 2005; Mojica et al. 2005). As bacteria encounter foreign DNA, a small portion of it, the proto-spacer, is sampled and incorporated between the palindromic repeats at the end of the CRISPR array (Barrangou et al. 2007; Kunin et al. 2007; Sorek et al. 2008). Upon re-exposure to the same foreign DNA, the host bacterium is able to degrade the DNA through a mechanism involving nucleic acid interference (Datsenko et al. 2012).

Two types of CRISPR arrays have been reported in prokaryotes: those associated with functional genes (CRISPR-Cas), and those without functional genes (Orphan CRISPR). Various species-specific CRISPR-Cas systems have been identified (Makarova et al. 2011). Recently, three CRISPR arrays, two with associated functional genes (CRISPR1-Cas, CRISPR3-Cas; both type II (Makarova et al. 2011) but distinguished by repeat sequence and Cas gene content) and one Orphan CRISPR (CRISPR2) were

identified in human pathogenic *Enterococcus faecalis* and *E. faecium* isolates (Palmer et al. 2010; Palmer and Gilmore 2010). Orphan CRISPR was ubiquitous in the *E. faecalis* strains studied whereas CRISPR-Cas were present only in strains that lacked anti-microbial resistance. A correlation between CRISPR-Cas presence and the absence of anti-microbial resistance and other virulence factors has also been observed in *E. faecalis* and *E. faecium* isolates of human clinical, food and environmental sources (Burley and Sedgley 2012; Lindenstrauss et al. 2011). Limited information about CRISPR from non-pathogenic enterococci is available, besides their identification and annotation in individual genomes (Grissa et al. 2007a, b). CRISPR from *E. hirae*, a common commensal species in the intestinal microbiota of pigs and chickens (Farrow and Collins 1985), and an occasional human pathogen (Canalejo et al. 2008; Chan et al. 2012), have not been characterized except for the CRISPR locus described in CRISPRdb for the *E. hirae* type strain ATCC 9790 (Grissa et al. 2007a, b).

The potential of CRISPR arrays as high-resolution strain-typing targets has led to intense study of their diversity and distribution in pathogenic bacteria (Brudey et al. 2006; Crawford 2003; Delannoy et al. 2012a; Delannoy et al. 2012b; Hauck et al. 2012; Pourcel et al. 2005). Studies of commensal organisms are less common and have tended to involve analysis of metagenomic sequence data rather than characterization of individual isolates from a community (Pride et al. 2011; Rho et al. 2012). A disadvantage of this approach is that despite use of amplification methods designed to target a narrow range of taxa, it is not possible to determine the species of origin for the CRISPR sequences identified (Pride et al. 2011). As a result, our understanding of the distribution

of CRISPR spacer sequences among more or less related bacterial taxa in natural microbial communities associated with humans and other animals is limited.

Enterococci are a major constituent of the fecal microbiome of young pigs, and although their abundance wanes post-weaning, they persist in the community as the microbiota adjust to a complex adult diet. We have shown previously the dynamic nature of the *Enterococcus* community in growing pigs, and described succession at the species and sub-species level (Vermette et al. 2010). The objective of this study was to determine prevalence and distribution of CRISPR arrays in cultured isolates of *E. faecalis* and *E. hirae* from a group of healthy pigs sampled at three ages, and to characterize their spacer and repeat sequences.

4.3 Materials and Methods

4.3.1 Bacterial Strains

Representative *Enterococcus* isolates (n=195) were selected from a large collection of pig fecal isolates (Vermette et al. 2010) based on their unique *cpn*60 UT sequences. Isolates are named according to pig number-age of the pig (in weeks)-isolate number (e.g. 81-9-A09). All isolates were grown in tryptone soy broth at 35ºC for 48 hours.

4.3.2 Genomic DNA Isolation

Genomic DNA was extracted from isolates using a modified salting out procedure (Martin-Platero et al. 2007), and stored at -20ºC in TE buffer (1 mM EDTA, 10 mM Tris-HCl, pH 7.4). The quality of DNA extractions was assessed by spectrophotometric measurements at 260 and 280 nm, and samples with an A_{260}/A_{280} ratio of 1.8-2.0 were

used for analysis. DNA integrity was further confirmed by successful amplification of the *cpn*60 UT region from each DNA extract using primers JH0064 and JH0065 as described previously (Vermette et al. 2010).

4.3.3 PCR amplification and sequencing of CRISPR arrays

Five sets of primers (A-E, Table 4.1) originally designed to detect CRISPR arrays in *E. faecalis* were used to detect the presence or absence CRISPR arrays (Palmer and Gilmore 2010).

Table 4.1 Primers for detection of CRISPR arrays in *Enterococcus*

Primer set[1]	Sequence (5´-3´)	Target[1]	Predicted product size	Annealing temp (°C)
A - For	CAGAAGACTATCAGTTGGTG	CRISPR1-Cas flanking region	315 bp (without CRISPR)	64.1
A - Rev	CCTTCTAAATCTTCTTCATAG			
B - For	GCGATGTTAGCTGATACAAC	CRISPR1-Cas *csn*1	783 bp	62.9
B - Rev	CGAATATGCCTGTGGTGAAA			
C - For	CTGGCTCGCTGTTACAGCT	Orphan CRISPR flanking region	Variable (with CRISPR)	63.5
C - Rev	GCCAATGTTACAATATCAAACA			
D - For	GCTGAATCTGTGAAGTTACTC	CRISPR3-Cas flanking region	224 bp (without CRISPR)	58.0
D - Rev	CTGTTTTGTTCACCGTTGGAT			
E - For	GATCACTAGGTTCAGTTATTTC	CRISPR3-Cas *csn*1	258 bp	64.1
E - Rev	CATCGATTCATTATTCCTCCAA			

[1] Palmer and Gilmore (2010)

Each 50 µl reaction mixture contained 1 × PCR reaction buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 2.5 mM $MgCl_2$, 50 pmol of each primer (Table 4.1), 0.2 mM dNTPs (Invitrogen, Burlington, ON), and 2.5 units Taq polymerase (Quanta Biosciences, Gaithersburg, MD). PCR amplifications were done in an Eppendorf Mastercycler EP gradient thermocycler with an initial denaturation step (95ºC for 3 minutes), followed by 40 cycles of denaturation (95ºC, 30 seconds), annealing (temperature in Table 4.1, 30 seconds) and extension (72°C, 1 minute), and a single final extension step (72ºC, 10 minutes). Plasmids containing PCR products for each CRISPR-related target from positive isolates were constructed for use as a PCR positive control for each primer set. Results of PCR reactions were evaluated by resolution of the products on 1.5% agarose gels, stained with ethidium bromide, and visualized and photographed under UV light. Amplified products were purified using EZ-10 spin column PCR purification kit (Bio Basic Inc., Markham, ON) and sequenced using their respective PCR primers (Table 4.1). Orphan CRISPR spacers and repeats were identified using CRISPRfinder (Grissa et al. 2007a, b), and predicted spacer sequences were identified by BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi) comparison to the Genbank non-redundant (nr) nucleotide database.

4.3.4 Genome walking

The GenomeWalker Universal Kit (ClonTech Laboratories, Inc., Mountain View, CA) was used to demonstrate genome duplication. Genome walking was performed according to the manufacturer's instructions. Four aliquots of genomic DNA from *E. hirae* isolate 71-9-A02 were digested with one of the restriction enzymes DraI, EcoRI, PvuII or StuI to create four adaptor-ligated genome walking libraries. PCR and

sequencing of genome walking libraries was carried out with the Genome Walking sequencing primer (ClonTech Laboratories, Inc., Mountain View, CA), primers listed in Table 4.1, and the primer JH0359 (5'-AAC GAT TCC GTC GAC GAC TAA TTC TTC-3') as appropriate.

4.4 Results and Discussion

4.4.1 CRISPR prevalence

The genomic locations of three CRISPR arrays in *E. faecalis* have been determined previously (Palmer et al. 2010; Palmer and Gilmore 2010), and five PCR primer sets were designed to locate and confirm the presence of each CRISPR (Table 4.1). Primer sets B and E target a region within the CRISPR-Cas gene *csn*1 (*cas*9 according to the nomenclature proposed by Makarova *et al*. (Makarova et al. 2011)), from the two type II CRISPR-Cas arrays designated by the authors as CRISPR1-Cas and CRISPR3-Cas, respectively. Products generated from these primers (783 bp expected for primer set B and 224 bp for primer set E) confirm the presence of the CRISPR-Cas *csn*1 gene, but give no information as to the genomic location of the CRISPR array. Primer sets A, C and D were designed to amplify sequences between *E. faecalis* genes directly up- and down-stream of CRISPR-Cas and Orphan CRISPR positions. Short PCR products are generated for primer set A (315 bp, targeting CRISPR1-Cas flanking region) or primer set D (224 bp, targeting CRISPR3-Cas flanking region) when there is no CRISPR-Cas array at that location. Primer set C amplifies the region spanning the insertion site for the Orphan CRISPR, resulting in variable product lengths depending on the numbers of repeats and spacers in the array. Sequences of all PCR products generated by primer set C were determined to identify repeat and spacer sequences.

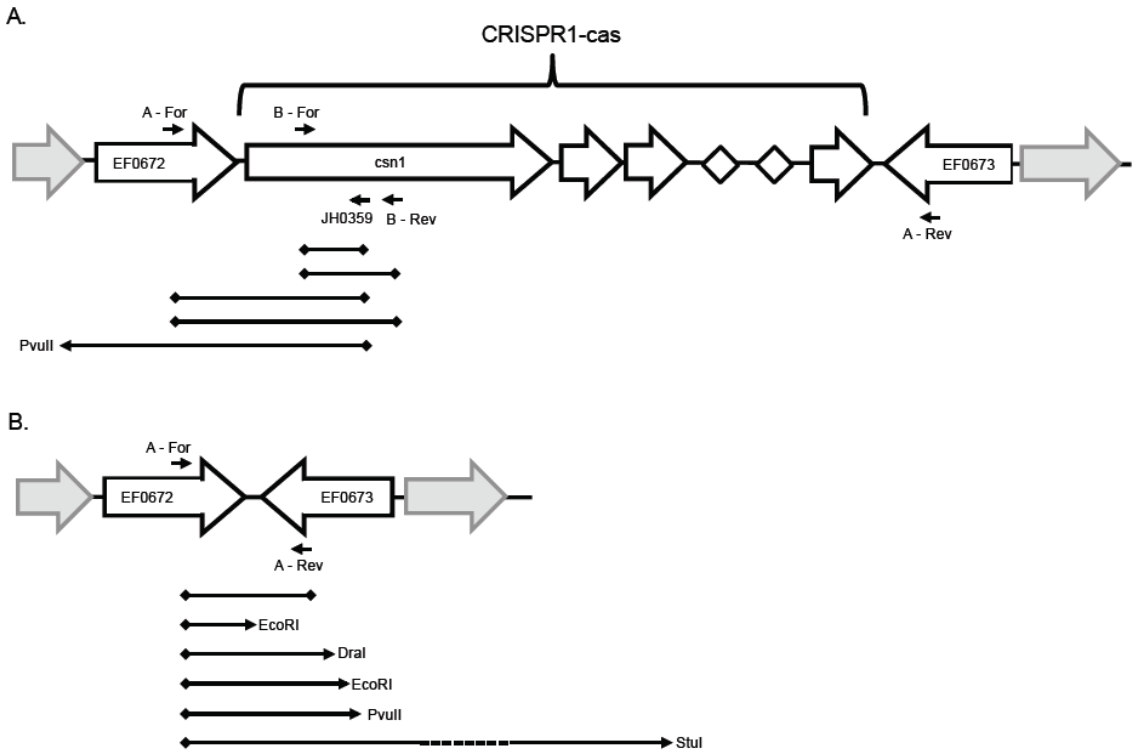
The prevalence of CRISPR-Cas and Orphan CRISPR arrays is shown in Table 4.2. CRISPR3-Cas, a novel CRISPR-Cas locus identified in *E. faecalis* strains Fly1, F1, RMC5 and T11 (Palmer et al. 2010; Palmer and Gilmore 2010) was not detected in any isolate, and all isolates were positive for the CRISPR3-Cas "missing from the genome" flanking region amplified by primer set D. Since all CRISPR-Cas arrays detected corresponded to the type II array previously designated CRISPR1-Cas, we will refer to these arrays using the preferred nomenclature CRISPR-Cas (Makarova et al. 2011) in the remainder of the manuscript.

Table 4.2 Prevalence of CRISPR arrays in commensal *Enterococcus* isolates from pig feces (n=195) based on PCR screening in this study

	Number(%) of isolates	
	E. faecalis (n=88)	*E. hirae* (n=107)
CRISPR1-Cas only	2(2)	14(13)
Orphan CRISPR only	40(45)	22(21)
CRISPR1-Cas and Orphan CRISPR	44(50)	31(29)
no CRISPR	2(2)	40(37)

Similar proportions of *E. faecalis* (52%, 46/88) and *E. hirae* (42%, 45/107) isolates were found to contain CRISPR-Cas, but prevalence of Orphan CRISPR arrays was substantially lower in *E. hirae* (49.5%, 53/107) than *E. faecalis* (95%, 84/88). The possibilities that the prevalence of Orphan CRISPR arrays in *E. hirae* was underestimated due to the application of primers that were designed based on *E. faecalis* sequences, and/or that a second insertion site not targeted by these primers exists, cannot be ruled out. However, the results do confirm the presence of Orphan CRISPR in a homologous position in both species, albeit in opposite orientations relative to the flanking ORFs (see below). Ongoing genome sequencing efforts focused on this culture collection will resolve this question.

For CRISPR-Cas, each isolate was expected to produce a product either with primer set A (indicating that no CRISPR-Cas array was present between homologues of *E. faecalis* V583 (NC_004668) genes EF0672 and EF0673) or primer set B (*csn*1 gene from CRISPR-Cas). However, 50% (44/88) of *E. faecalis* and 40% (43/107) of *E. hirae* isolates generated PCR products with both primer sets. Several isolates were retested from freshly grown single colonies to ensure they were pure cultures, with no difference in the PCR result. This left two possible explanations: either the CRISPR-Cas array in these isolates was located in a different genomic position than previously reported, or there was a genome duplication including the CRISPR-Cas insertion site, with only one site occupied. To resolve this situation, *E. hirae* isolate 71-9-A02 was selected as a representative of the present/absent CRISPR-Cas isolates, and genome walking was used to determine the genomic position of the CRISPR-Cas *csn*1 gene (Supplementary figure 4.1).



Supplementary Figure 4.1 Genome arrangements in *E. hirae* isolate 71-09-A02 around CRISPR1-cas array based on genome walking results.
Predicted open reading frames flanking the predicted CRISPR1-cas location are labeled according to their homologues in *E. faecalis* V583 (Genbank Accession NC_004668). Lines represent PCR products generated from indicated primers (diamond ends) or genome walking restriction sites (arrows and indicated restriction enzyme). Primer JH0359 was designed to work with the genome walking adaptor primer to amplify the region upstream of the CRISPR1-cas locus from PvuII, EcoRI, DraI or StuI genome walking libraries. From the first "walk", PCR with the PvuII library generated a product (1A) that included the CRISPR1-cas csn1 gene downstream of a homologue of the E. faecalis gene EF0672. Since this was the reported gene upstream of the CRISPR1-cas locus, we combined the forward primer from Primer set A with either the reverse primer from Primer set B or JH0359 and generated the PCR products consistent with the CRISPR1-cas array being in the expected position (1A). A second genome "walk" was then done starting within the EF0672 homologue using the Primer set A forward primer and the same genomic DNA libraries as before. This time, all four libraries generated products (EcoRI generated two products) that were determined to include the end of the EF0672 homologue, a short intergenic region, followed by a homologue of the EF0673 gene (1B).

Two arrangements, with and without CRISPR-Cas, were detected within the genome of this isolate, consistent with a duplication of the genomic region surrounding the EF0672 homologue, with only one of the duplicates containing a CRISPR-Cas array. This was perhaps not surprising since *Enterococcus* genomes are known to be highly plastic and genome rearrangements are common (Aakra et al. 2007; Lepage et al. 2006). However, our search for multiple occurrences of the predicted ORFs flanking the CRISPR-Cas locus in other complete *Enterococcus* genomes yielded no evidence of a similar duplication.

4.4.2 CRISPR Annotation

Orphan CRISPR PCR products produced with primer set C were sequenced and high quality sequence data was obtained for 123/137 isolates (81 *E. faecalis* isolates and 42 *E. hirae* isolates). Array sequences were annotated using CRISPRfinder (Grissa et al. 2007a, b). The most common repeat sequence for each species (identified in ≥90% of Orphan CRISPR elements examined) was selected for spacer annotation. Spacer sequences that differed by at least 2 nucleotides were considered different.

Distinct repeat sequences were identified for *E. faecalis* and *E. hirae* using CRISPRfinder. Species-specificity of repeat sequences has previously been shown for lactic acid bacteria in a range of genera including *Enterococcus* (Horvath et al. 2009), and the repeat sequence identified for *E. faecalis* matches the repeat sequence reported for Orphan CRISPR in human clinical *E. faecalis* isolates (Palmer et al. 2010; Palmer and Gilmore 2010). Interestingly, while the direct repeat sequences within Orphan CRISPR arrays in *E. faecalis* were remarkably consistent (5'-GTT TTA GAG TCA TGT TGT TTA GAA TGG TAC CAA AAC-3', with a variation of no more than two nucleotides), a

different pattern was apparent in *E. hirae*. In *E. hirae*, the repeats flanking the terminal spacer in each array (5'-GTT TTA GAG CCA TGT TGT TAA AAA ACA AAC TAT CAC-3') were distinct from the consensus *E. faecalis* repeat, while the remaining repeats in these arrays were almost identical to the consensus *E. faecalis* repeat. The *E. hirae*-specific repeat sequence did not match the repeat sequence for *E. hirae* ATCC 9790 described in CRISPRdb. The number of repeats in individual isolates ranged from two to fifteen for both *E. faecalis* and *E. hirae*.

A total of 692 spacers were identified in the 123 sequenced Orphan CRISPR arrays, representing 42 unique spacer sequences (Supplementary Table 4.1), and 24 unique spacer patterns (17 in *E. faecalis*, 7 in *E. hirae*) (Table 4.3). The number of spacer elements ranged from 1 to 14 in *E. faecalis* arrays (average 5), and 1 to 14 in *E. hirae* arrays (average 7). All spacers were 30 or 31 bp in length.

Supplementary Table 4.1 Spacer sequences and presence/absence in *E. faecalis* and *E. hirae*

Spacer ID	Spacer sequence	*E. faecalis*	*E. hirae*
1	ATGATGAAATTAATCGAATTAGACAACGGA	+	+
2	TCAGCACATGGATTATGATATTGTCAAAGA	+	+
3	TTGCCAATGGTTGACCTTATCAACTTACAC	+	+
4	TTATCGATTTCTTGTTGGATCTGAGTACTT	+	+
5	TGAGGTTTGTGTAACCAATTTACTGAAGCT	+	+
6	TGCACCGTCAACGTATGTTGGGAGCTCGTA	+	+
7	AGCAAAAGAAGAAAAGAAAACAAAAGTTGC	+	+
8	ATGTTGAAGTAGAGCGTGATGCAATCCTAA	+	+
9	GCTAATTTAAAGGCAAAGGCAAGAATAGAA	+	+
10	TGGTGAAGATGTTGATTCTGGATTATCAAA	+	+
11	CACTTCCCAAATAGAAAGGACGATGAAACA	+	+
12	CTAATGTCAAAACAGCAGCTACATTTCTCC	+	+
13	TGGGTTGACTAAAGAGCCGTCAAAAGTTTT	+	+
14	TCAAGAAATTGCATTAAGTTCAAAAAATTT	+	-
15	GCTGTAGCGTGTGTTTCCCCATAGTCATTC	+	+
16	AGCTGAAAGAGCAGGACCAGAAAGGCCCGC	+	+
17	GTTTCGGCAAGGCGAACCCTGGTATTTGGT	+	+
18	AACAATTAGCGCCAGCTGGCCAGATTCAAT	+	-
19	ACAGAAATATCAAACGCAAAAAAAGGTTTC	+	-
20	GACTTACAAAAGACTGTGATTTACGTTATA	+	-
21	AAACTTTTTTGATTTGGCTTTTTCTCCCCT	+	-
22	CGGCAAAGGAGGAAAAGAAAACAAAAGTTG	+	-
23	GGGATTTACAATAACTTCGCTACCATCTTC	+	-
24	TTATGTGATTGAGGGAATTTTGATTGATGC	+	-
25	ATAAGGCTAAAGAATCTATCAAGAAAGTAA	+	+
26	CTTAAAGCGTTCCGAATAGCGTTCTAAGAA	+	-
27	TAAGCATTTTTTTGATGTGTAATACTTATC	+	-
28	CTTCCAAAATTTAGAACAAGNAAAAAAAGAC	+	-
29	GCATTTTTGTAGTTCTTACTATTCTTGCTT	+	-
30	TCGACCGTCCCATGTTGGAGCTCCAATCAA	+	-
31	CCTTATTTAAAAGACGGCAATATTAAAAACA	+	-
32	AAAAAAAACTTTGCAATTAGTTAGATCTAA	+	-
33	AAAAAACTAGCTTTAAAGTATCACCCAGAC	+	-
34	AAGAAGCAAAAGAGAACATGATTGCCGTTT	+	-
35	ATTTATATTCAATCACAGTAGCAATACCTT	+	-
36	TGATCGTGGATTTTCTGACAGGTTGGATCA	+	-
37	CCAAGTTGGTACGATTACTCAAAAATTAGA	+	-
38	GCCAAAGGAGGAAAGAAACCAAAAGTTGCC	+	-
39	CGCCAAAAGAAGAAAAGAAAACAAAAGTTG	+	-
40	GGCAAAGGAGGAAAGGAAACCAAAAGTTGC	+	-
41	AAATTTTTATCACTTAATGCAATTTCTTGA	-	+
42	GCAACTTTTGTTTTCTTTTCTTCTTTGGCC	-	+

Table 4.3 Prevalence of Orphan CRISPR unique spacer patterns among isolates of *E. faecalis* and *E. hirae* collected from pigs at 3, 9, and 15 weeks of age

	No. of isolates (No. of pigs)		
Spacer pattern[1]	**3 wks (n=61)**	**9 wks (n=22)**	**15 wks (n=40)**
E. faecalis			
15.16.*14*.	44(4)	0	0
17.	11(3)	0	0
1.2.3.4.5.6.7.8.9.10.11.12.13.*14*.	0	3(1)	6(4)
23.24.30.31.21.	3(1)	0	0
1.2.3.4.	0	2(1)	0
1.2.3.4.5.6.7.	0	1(1)	0
1.4.5.6.7.8.9.10.11.12.13.*14*.	0	0	1(1)
12.13.*14*.	0	0	1(1)
18.19.20.21.18.19.20.21.	1(1)	0	0
22.8.9.10.11.12.13.*14*.	0	0	1(1)
23.24.25.26.27.28.29.30.31.21.	0	0	1(1)
32.3.9.11.12.13.*14*.	0	1(1)	0
33.34.35.36.37.	0	1(1)	0
5.6.22.8.9.10.11.12.13.*14*.	0	0	1(1)
6.38.8.9.10.11.12.13.*14*.	0	0	1(1)
6.39.8.9.10.11.12.13.*14*.	0	0	1(1)
6.40.8.9.10.11.12.13.*14*.	0	0	1(1)
E. hirae			
5.4.3.2.1.	0	1(1)	0
41.13.12.11.10.9.8.7.6.5.4.3.2.1.	0	2(1)	13(3)
41.16.15.	1(1)	10(5)	10(2)
41.16.	0	0	1(1)
41.13.12.11.10.9.8.42.	0	0	2(1)
41.13.12.11.10.9.8.25.6.	0	1(1)	0
17.	1(1)	0	0

[1]Spacers are numbered according to Supplementary table 4.1 and presented in the order encountered in the genome from forward to reverse primer landing site (primer set C, Table 4.1). Putative ancestral spacers are in italics.

4.4.3 Spacer Identification

Spacer sequences were compared to the NCBI non-redundant (nr) database for identification using BLASTn. Hits with ≥90% identity over ≥90% of the query length were considered significant. Only 7 of the 42 spacers could be identified as previously described mobile elements (Table 4.4). Three spacers matched *Enterococcus*-specific phages (SAP6, phiFL3B, and phiFL3A), three more spacers matched *E. faecalis*-specific plasmids and one spacer matched part of an *Enterococcus*-specific pathogenicity island (a mobile 150 kb fragment that carries several virulence determinant (Coburn et al. 2007)).

In addition to the Genbank nr database, spacer sequences were compared to two published pig virome data sets (NCBI Sequence Read Archive ID SRA030664 and SRA045429) (Allen et al. 2011; Mafamane et al. 2011; Shan et al. 2011; Vorontsova et al. 2015). All spacers from *E. faecalis* and *E. hirae* were compared to over 2 million sequence reads from the pig virome using BLASTn and the same criteria (≥90% identity over ≥90% query coverage). Only spacer 42 (unique to *E. hirae* isolates) and spacer 29 (unique to *E. faecalis* isolates) matched sequences from the pig virome data (gnl|SRA|SRR358776.61063.2 and gnl|SRA|SRR358789.109289.2, 92% and 93% identity, respectively). These spacers did not match any sequence in the Genbank nr database (Table 4.4).

Table 4.4 Spacer identification in *E. faecalis* and *E. hirae*[1, 2]

Spacer	Description[1,2]	Query coverage (%)	Identity (%)
1	*Enterococcus* phage SAP6	100	100
6	*Enterococcus* phage SAP6	93	100
8	*Enterococcus* phage phiFL3B, phiFL3A	100	97
12	*Enterococcus faecalis* plasmid pLG2 transposase (pLG2-0010)	93	100
20	*Enterococcus faecalis* plasmid EF62pB, pBee99, pMG2200, pyl14, pCF10, pTEF2	100	100
27	*Enterococcus faecalis* plasmid EF62pB, pCF10, pTEF2	100	100
29	Pig virome	93	93
33	*Enterococcus faecalis* pathogenicity island	100	94
42	Pig virome	90	92
2, 4, 5, 7, 9, 11, 13-16, 18, 21, 23, 24, 31, 34, 41	*Enterococcus* chromosome	96-100	100
3, 10, 17, 19, 22, 25, 26, 28, 30, 32, 35-40	Unidentified		

[1]Only the best match is reported for each spacer sequence, unless a tie was reported
[2]Only hits with ≥90% identity were included

Seventeen spacer sequences matched *Enterococcus* chromosomal sequences, including both sequences within predicted open reading frames and non-annotated, or intergenic regions (Table 4.4). It has been suggested that spacers with similarity to chromosomal sequences from other closely related organisms are the result of exposure to DNA from other bacteria in the community through conjugation, transduction, or transformation (Brodt et al. 2011). Alternatively, some of these sequences may correspond to unannotated spacers within the target genomes.

Sixteen of 42 (38%) spacer sequences had no significant matches to any sequences in Genbank nr or the pig virome data sets, indicating a lack of information related to mobile elements that affect the pig fecal microbiota.

Protospacer adjacent motifs (PAMs) are di- or tri- nucleotide sequences present within 1-4 nucleotides on either side of the protospacer, and are thought to play a critical role in recognition and acquisition of spacers (Mojica et al. 2009). To identify potential PAMs, we retrieved sequences identical to spacers from plasmids and phage described in Table 4.4. In 5/7 cases, a GG dinucleotide was found immediately adjacent to the spacer sequence. Jinek et al. (2012) have shown recently that in the type II CRISPR-Cas system (present in isolates from our study), protospacer recognition by Cas9 requires a GG dinucleotide-containing PAM sequence, suggesting that a type II CRISPR-Cas system inserted spacers in these Orphan CRISPR arrays.

4.4.4 Spacer Distribution

Unique spacer patterns detected are shown in Table 4.3, with the order of the spacers presented as they were encountered in the genome moving from the forward to reverse primer landing sites for primer set C (Table 4.1). It was immediately apparent that

the *E. hirae* arrays are in the opposite orientation to the *E. faecalis* arrays. Spacer 14 occurred most frequently as the terminal (oldest) spacer in *E. faecalis* Orphan CRISPR arrays, and spacer 41 in the corresponding position in *E. hirae* arrays. These spacers are putative "ancestral spacers" for these species; a suggestion further supported by the fact that spacer 14 occurs as the terminal spacer in Orphan CRISPR arrays annotated in *E. faecalis* strains OG1RF and D32 (derivative of human clinical isolate OG1, and Danish pig isolate from 2001, respectively). Both of these strains also include CRISPR-Cas, which do not contain spacer 14, consistent with CRISPR-Cas actively incorporating new spacers and ejecting older ones as the array increases in size. Spacer 41, the proposed ancestral spacer for the *E. hirae* isolates in our study was not identified as a spacer in any published *Enterococcus* genomes including *E. hirae* ATCC 9790, although it was identical to sequence in *E. faecalis* strains 7L76 (intergenic region between two predicted ORFs) and V583 (within a predicted ORF of unknown function).

Identical spacers and spacer patterns were detected among isolates from different individual pigs (Supplementary Tables 4.2, Supplementary Tables 4.3 and Table 4.3), which is perhaps not surprising given that these animals were raised in the same barn, and were in some cases litter-mates (Vermette et al. 2010). Six of the 24 unique spacer patterns were detected in at least two pigs, and one of the most prevalent patterns (*E. hirae* spacer pattern 15.16.41) was detected in 21 isolates from 8 different pigs at 3, 9 and 15 weeks of age. This is in stark contrast to observations made in humans where person-specific spacer populations are resolvable, reflecting the individuality of the hosts and their environments. Rho *et al*. (Rho et al. 2012) extended this observation, reporting that spacer populations were distinct even between different body sites on individual subjects.

Supplementary Table 4.2 Prevalence and distribution of spacers 1-42 in Enterococcus isolates. Presence and absence are indicated by 1 and 0, respectively.

		Spacers																																														
Species	**Isolate_ID**	**1**	2	3	4	5	**6**	7	8	9	10	11	**12**	13	14	15	16	17	**18**	19	20	21	22	23	**24**	25	26	27	28	29	**30**	31	32	33	34	35	**36**	37	38	39	40	41	**42**	**Total**
faecalis	88_3_A01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	1
faecalis	81_9_A02	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	81_9_A11	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	81_9_H09	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	57_15_A08	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	68_15_F01	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	78_15_A02	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	78_15_A04	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	88_15_G10	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	81_15_E07	0	0	0	0	0	**0**	0	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	8
faecalis	81_3_A03	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_15_E02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0**	0	0	1	0	1	**1**	1	1	1	1	1	**1**	1	0	0	0	0	**0**	0	0	0	0	0	**0**	10
faecalis	88_15_B01	**0**	0	0	0	0	**1**	0	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	0	0	0	**0**	9
faecalis	88_15_C07	**0**	0	0	0	0	**1**	0	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	1	0	0	**0**	9
faecalis	88_15_H09	**0**	0	0	0	0	**1**	0	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	1	0	**0**	9
faecalis	81_15_E03	**0**	0	0	0	0	1	**1**	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	10
faecalis	81_9_B06	**1**	1	1	1	1	**1**	1	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	7
faecalis	81_9_D11	**0**	0	1	0	0	**0**	0	0	1	0	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	0	0	0	**0**	0	0	0	0	0	**0**	7
faecalis	88_3_H07	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	1	0	1	**1**	0	0	0	0	0	**1**	1	0	0	0	0	**0**	0	0	0	0	0	**0**	5
faecalis	88_3_H08	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	1	0	1	**1**	0	0	0	0	0	**1**	1	0	0	0	0	**0**	0	0	0	0	0	**0**	5
faecalis	81_9_F12	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	1	1	1	**1**	1	0	0	0	0	**0**	5
faecalis	88_15_A05	**1**	1	1	1	1	**1**	1	1	1	1	1	**1**	1	1	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	14
faecalis	71_3_F08	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**1**	1	1	1	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	4
faecalis	78_9_A10	**1**	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	4
faecalis	78_9_D04	**1**	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	4
faecalis	57_3_A01	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3

		Spacers																																									
Species	**Isolate_ID**	**1**	2	3	4	5	**6**	7	8	9	10	11	**12**	13	14	15	16	17	**18**	19	20	21	22	23	**24**	25	26	27	28	29	**30**	31	32	33	34	35	**36**	37	38	39	40	41	**42**	**Total**
faecalis	57_3_G01	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	57_3_G09	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_F06	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_A01	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_A04	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_A12	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_B02	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_C08	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_F03	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_F07	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_F08	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_G10	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	78_3_H08	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_A01	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_A02	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_A09	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_B01	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_B04	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_D05	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_E02	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	81_3_G04	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	88_3_D05	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	88_3_D08	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	88_3_E09	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	88_3_F07	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	88_3_G07	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3
faecalis	88_3_H11	**0**	0	0	0	0	**0**	0	0	0	0	0	**0**	0	1	1	1	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	0	0	0	0	0	**0**	3

		Spacers																																									
Species	**Isolate_ID**	**1**	2	3	4	5	**6**	7	8	9	10	11	**12**	13	14	15	16	17	**18**	19	20	21	22	23	**24**	25	26	27	28	29	**30**	31	32	33	34	35	**36**	37	38	39	40	41	**42**	**Total**
faecalis	88_3_A03	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_A07	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_A08	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_A10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_A11	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_B07	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_C02	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_C09	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_C10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_D01	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_D04	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_E12	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_F09	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	88_3_G01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	5
faecalis	88_3_H10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	81_15_G08	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
faecalis	88_15_A12	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
faecalis	68_3_A02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	68_3_A06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	68_3_B01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	68_3_C05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	68_3_D03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	68_3_D09	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	68_3_F02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	68_3_G05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	78_3_A06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
faecalis	88_3_A02	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

Species	Isolate_ID	Spacers																																									
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	Total
faecalis	68_9_A12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
hirae	57_3_H11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	67_3_C05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
hirae	67_9_A02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	68_9_C04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	78_9_A02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	78_9_A05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	78_9_B08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	78_9_C01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	78_9_A01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	81_9_A05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_9_B02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	78_9_C03	0	0	0	0	0	1	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9
hirae	68_9_E03	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
hirae	67_9_D07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	81_9_G12	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	81_9_B05	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	78_15_A01	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_A02	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_A01	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_A03	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_A09	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_A10	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_B06	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_B12	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_C01	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_C04	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14

		Spacers																																										
Species	**Isolate_ID**	**1**	2	3	4	5	**6**	7	8	9	10	11	**12**	13	14	15	16	17	**18**	19	20	21	22	23	**24**	25	26	27	28	29	**30**	31	32	33	34	35	**36**	37	38	39	40	41	**42**	**Total**
hirae	88_15_F11	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	81_15_E01	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	8
hirae	81_15_E05	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	8
hirae	81_15_E04	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	71_15_G08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	78_15_C06	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	14
hirae	88_15_D10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_D06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_D08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_H06	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_B04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_D01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_D03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_F12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_H03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
hirae	88_15_H01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2

Supplementary Table 4.3 Orphan CRISPR array spacer patterns for all isolates.
Patterns are presented according to their orientation in the genome, relative to the flanking ORFs

Species	Isolate	Pig	Age (wks)	Pattern
faecalis	68-3-A01	68	3	17.
faecalis	68-3-A02	68	3	17.
faecalis	68-3-A06	68	3	17.
faecalis	68-3-B01	68	3	17.
faecalis	68-3-C05	68	3	17.
faecalis	68-3-D03	68	3	17.
faecalis	68-3-D09	68	3	17.
faecalis	68-3-F02	68	3	17.
faecalis	68-3-G05	68	3	17.
faecalis	78-3-A06	78	3	17.
faecalis	88-3-A01	88	3	17.
faecalis	78-9-A10	78	9	1.2.3.4.
faecalis	78-9-D04	78	9	1.2.3.4.
faecalis	81-9-B06	81	9	1.2.3.4.5.6.7.
faecalis	81-9-A02	81	9	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	81-9-A11	81	9	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	81-9-H09	81	9	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	57-15-A08	57	15	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	68-15-F1	68	15	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	78-15-A2	78	15	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	78-15-A4	78	15	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	88-15-A05	88	15	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	88-15-G10	88	15	1.2.3.4.5.6.7.8.9.10.11.12.13.14.
faecalis	81-15-G8	81	15	1.4.5.6.7.8.9.10.11.12.13.14.
faecalis	88-15-A12	88	15	12.13.14.
faecalis	57-3-A01	57	3	15.16.14.
faecalis	57-3-G01	57	3	15.16.14.
faecalis	57-3-G09	57	3	15.16.14.
faecalis	78-3-A01	78	3	15.16.14.
faecalis	78-3-A04	78	3	15.16.14.
faecalis	78-3-A12	78	3	15.16.14.
faecalis	78-3-B02	78	3	15.16.14.

Species	Isolate	Pig	Age (wks)	Pattern
faecalis	78-3-C08	78	3	15.16.14.
faecalis	78-3-F03	78	3	15.16.14.
faecalis	78-3-F07	78	3	15.16.14.
faecalis	78-3-F08	78	3	15.16.14.
faecalis	78-3-F6	78	3	15.16.14.
faecalis	78-3-G10	78	3	15.16.14.
faecalis	78-3-H08	78	3	15.16.14.
faecalis	81-3-A01	81	3	15.16.14.
faecalis	81-3-A02	81	3	15.16.14.
faecalis	81-3-A09	81	3	15.16.14.
faecalis	81-3-A3	81	3	15.16.14.
faecalis	81-3-B01	81	3	15.16.14.
faecalis	81-3-B04	81	3	15.16.14.
faecalis	81-3-D05	81	3	15.16.14.
faecalis	81-3-E02	81	3	15.16.14.
faecalis	81-3-G04	81	3	15.16.14.
faecalis	88-3-A02	88	3	15.16.14.
faecalis	88-3-A03	88	3	15.16.14.
faecalis	88-3-A07	88	3	15.16.14.
faecalis	88-3-A08	88	3	15.16.14.
faecalis	88-3-A10	88	3	15.16.14.
faecalis	88-3-A11	88	3	15.16.14.
faecalis	88-3-B07	88	3	15.16.14.
faecalis	88-3-C02	88	3	15.16.14.
faecalis	88-3-C09	88	3	15.16.14.
faecalis	88-3-C10	88	3	15.16.14.
faecalis	88-3-D01	88	3	15.16.14.
faecalis	88-3-D04	88	3	15.16.14.
faecalis	88-3-D5	88	3	15.16.14.
faecalis	88-3-D8	88	3	15.16.14.
faecalis	88-3-E12	88	3	15.16.14.
faecalis	88-3-E9	88	3	15.16.14.
faecalis	88-3-F09	88	3	15.16.14.
faecalis	88-3-F7	88	3	15.16.14.
faecalis	88-3-G7	88	3	15.16.14.

Species	Isolate	Pig	Age (wks)	Pattern
faecalis	88-3-H10	88	3	15.16.14.
faecalis	88-3-H11	88	3	15.16.14.
faecalis	71-3-F08	71	3	18.19.20.21.18.19.20.21.
faecalis	81-15-E7	81	15	22.8.9.10.11.12.13.14.
faecalis	81-15-E02	81	15	23.24.25.26.27.28.29.30.31.21.
faecalis	88-3-G01	88	3	23.24.30.31.21.
faecalis	88-3-H07	88	3	23.24.30.31.21.
faecalis	88-3-H8	88	3	23.24.30.31.21.
faecalis	81-9-D11	81	9	32.3.9.11.12.13.14.
faecalis	81-9-F12	81	9	33.34.35.36.37.
faecalis	81-15-E3-	81	15	5.6.22.8.9.10.11.12.13.14.
faecalis	88-15-B1	88	15	6.38.8.9.10.11.12.13.14.
faecalis	88-15-C7	88	15	6.39.8.9.10.11.12.13.14.
faecalis	88-15-H9	88	15	6.40.8.9.10.11.12.13.14.
hirae	67-3-C05	67	3	17
hirae	68-9-E3	68	9	5.4.3.2.1.
hirae	81-9-B5	81	9	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	81-9-G12	81	9	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	81-15-E04	81	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-A01	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-A03	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-A09	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-A10	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-B06	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-B12	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-C01	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-C04	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-F11	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	78-15-C6	78	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	78-15-A1	78	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	88-15-A2	88	15	41.13.12.11.10.9.8.7.6.5.4.3.2.1.
hirae	57-3-H11	57	3	41.16.15.
hirae	67-9-A2	67	9	41.16.15.
hirae	67-9-D7	67	9	41.16.15.
hirae	68-9-C04	68	9	41.16.15.

Species	Isolate	Pig	Age (wks)	Pattern
hirae	78-09-A01	78	9	41.16.15.
hirae	78-9-A2	78	9	41.16.15.
hirae	78-9-A5	78	9	41.16.15.
hirae	78-9-B8	78	9	41.16.15.
hirae	78-9-C1	78	9	41.16.15.
hirae	81-9-A5	81	9	41.16.15.
hirae	88-9-B02	88	9	41.16.15.
hirae	71-15-G08	71	15	41.16.15.
hirae	88-15-B04	88	15	41.16.15.
hirae	88-15-D01	88	15	41.16.15.
hirae	88-15-D03	88	15	41.16.15.
hirae	88-15-F12	88	15	41.16.15.
hirae	88-15-H03	88	15	41.16.15.
hirae	88-15-D10	88	15	41.16.15.
hirae	88-15-D6	88	15	41.16.15.
hirae	88-15-D8	88	15	41.16.15.
hirae	88-15-H6	88	15	41.16.15.
hirae	88-15-H01	88	15	41.16.
hirae	81-15-E1	81	15	41.13.12.11.10.9.8.42.
hirae	81-15-E5	81	15	41.13.12.11.10.9.8.42.
hirae	78-9-C3	78	9	41.13.12.11.10.9.8.25.6.

In general, different CRISPR array spacer patterns were detected in individual pigs at different ages, reflecting the succession patterns of enterococci we have previously reported in this environment (Vermette et al. 2010). However, there were some examples where identical spacer patterns were detected in the same pig at different ages (Table 4.3), suggesting that some strains persisted in the intestine for periods of at least 6 weeks. No obvious examples of spacer succession (gain of new spacers and loss of existing spacers) were observed in Orphan CRISPR spacer arrays over sequential samplings of the *Enterococcus* community. However, our study focused on Orphan CRISPR (without Cas genes), which may be inactive, or updated by CRISPR-Cas elements working in *trans* (Bhaya et al. 2011; Klatt et al. 2011). The arrangement where both CRISPR-Cas and Orphan CRISPR are present in the genome was detected in 44/88 (50%) and 31/107 (29%) of *E. faecalis* and *E. hirae*, respectively (Table 4.2). Also, the time frame (12 weeks) over which these isolates were collected is relatively short in comparison to studies where CRISPR succession is apparent (Pride et al. 2011).

While the distribution of spacers and spacer arrangement among the pigs in this study reflects the "pan-microbiota" of a group of genetically related animals reared concurrently in a shared environment, the observation that only two spacers (one unique to each species) matched previously published pig virome sequences shows that this cohort of animals is distinct from others in terms of the strains comprising their microbiota, and mobile genetic elements to which they are exposed. Combined with observations in human systems, where CRISPR profiles are apparently more specific to individuals, this highlights one of the potential advantages to studying the intestinal microbial ecology in food animal models such as pigs.

Surprisingly, 40% (17/42) of the different spacer sequences detected in Orphan CRISPR arrays were shared between both species (Supplementary Tables S1 and S2). Evidence of spacers shared by different species of bacteria is scarce, but this phenomenon has been reported in Archaea where it reflects "networks of gene exchange" in communities (Brodt et al. 2011). Furthermore, it demonstrates that many of the agents and mechanisms represented by these spacers are not species-specific.

This is, to our knowledge, the first study of CRISPR diversity in cultured commensal enterococci isolates. We have shown that although *E. faecalis* and *E. hirae* have distinct repeat sequences, their shared spacer sequences reflect their common environment and the extensive networking and transfer of genetic material within the dynamic intestinal microbiota. Along with high-resolution phylogenetic targets such as the *cpn*60 gene, CRISPR sequences provide an additional molecular tool for detection and monitoring of species and strain-level succession in microbial communities.

CHAPTER 5 - Characterization of genomic diversity in *cpn*60 defined *Enterococcus hirae* ecotypes and relationship to competitive fitness

Fifth Chapter Transition

Results presented in the previous chapters confirm the ability of the *cpn*60 UT sequence to resolve species and sub-species level diversity within the genus *Enterococcus*. Results of the application of *cpn*60 UT sequencing, genome fingerprinting and CRISPR analysis of representative *Enterococcus* isolates from pig feces indicate that the subgroups identified likely represent ecotypes within a community. We observed genomic diversity within and between ecotypes using GTG rep PCR, which could be a result of genome re-arrangements and/or genome content differences. Although subspecies clustering of *E. faecalis* by *cpn*60 UT sequencing was not apparent in the GTG rep PCR analysis, the resolution of *E. hirae* subgroups was clear by either method. In the following chapter, whole genome sequencing of six representative isolates of three *E. hirae* subgroups was conducted to identify genomic features that define ecotypes within this species. Results were interpreted with the goal of explaining previously described differences in their relative abundance, and succession trends in feces of growing pigs. In addition, *in vitro* competition experiments were conducted to determine if genome content differences could predict fitness.

5.1 Abstract

The astounding complexity and dynamic nature of gut microbial communities limit the study of species and strain level genomic diversity. In a previous study, we demonstrated that phenotypically and genotypically distinct ecotypes of *Enterococcus* emerge over time in the pig fecal microbiome. In this study, we characterized genomic diversity in *E. hirae* ecotypes. Genome sequences for six representative isolates (two from each of three ecotypes) were compared. Differences in phosphotransferase systems, amino acid metabolism pathways for glutamine, proline and selenocystiene, potassium-transporting ATPases, copper homeostasis systems and putative prophage associated sequences, CRISPRs and antibiotic resistance genes were observed. Differences in the lac family phophotransferase system corresponded to previous observations regarding lactose utilization by the isolates. We tested fitness of the *E. hirae* ecotypes by conducting *in vitro* growth competition assays in pig fecal extract medium. Isolates from *E. hirae-1* and *E. hirae-2* were able to out-compete isolates from the *E. hirae-3* ecotype, which is consistent with the relatively low abundance of *E. hirae-3* relative to *E. hirae-1* and *E. hirae-2* previously observed in the pig fecal microbiome, and with observed gene content differences between the ecotypes related to biosynthetic capacity. Results of this study provide a genomic basis for the definition of ecotypes within *E. hirae* and contribute to an understanding of the *Enterococcus* community structure observed within the fecal microbiome of healthy pigs.

5.2 Introduction

Natural microbial communities undergo changes in composition and function under the influence of various factors such as temperature, pH and available nutrients, among others (Cariveau et al. 2014; Isaacson and Kim 2012; Morato et al. 2014; Ottesen et al. 2013; Torok et al. 2011; Tran et al. 2012). Characterization of diversity in bacterial community composition over time or with changes in environmental factors is important since community structure defines community function (Berga et al. 2012; Romani et al. 2014).

High throughput techniques have improved our understanding of highly inter-related metabolic, genetic, and biochemical networks that characterize complex microbial communities in their natural environments. Two important questions in microbial community ecology - Who is there? What are they doing? - have been answered by using multiple approaches (Morgan and Huttenhower 2014). A database of genes in a microbial community can be obtained by using next generation DNA sequencing methods (metagenomics), whereas phenotypic patterns can be studied by looking at mRNA expression profiles (transcriptomics), protein expression (proteomics), or metabolite production (metabolomics) (Faith et al. 2011; Wu et al. 2011a). However, marker gene targeted microbial community profiling studies remain the most common approach for determining community composition, and are often used to develop the first description of a novel microbial environment.

Studies in which changes in community composition at the species level are observed over time are rare. It has been suggested that discrete clusters known as ecotypes exist within bacterial species. Ecotypes emerge within bacterial species through

periodic selection, and can be discriminated on the basis of phenotypic, ecological and DNA sequence characteristics (Wiedenbeck and Cohan 2011). In phylogenetic studies based on conserved targets such as the 16S rRNA gene and the *cpn*60 UT, ecotypes are detected as clusters of sequences where the average distance between ecotypes is greater than the average distances within the cluster. In sequence based microbial community studies, this level of resolution may go undetected as sequences are aggregated into operational taxonomic units based on a sequence identity cutoff, or they may be dismissed as next-generation sequencing artifacts (Kunin et al. 2010; Schloss et al. 2011).

Substantial evidence for diversity among closely related strains within a species has been generated from genomic comparisons. Gene level diversity is a result of the rapid growth rate, mutations, acquisition and loss of genes and influence of environmental factors characteristic of complex bacterial communities. Doolittle and Zhaxybayeva (2009) suggested that the presence of unique genes may play a role in determining the ecological fitness of an isolate in a complex bacterial community, and that even closely related bacteria can occupy ecologically distinct niches. A few studies directed at identification and characterization of genomic heterogeneity in ecotypes have been conducted and have provided valuable information about genomic features that define ecotypes (Cohan 2006, 2009; Klatt et al. 2011; Kopac et al. 2014; Wiedenbeck and Cohan 2011). Genome content differences between members of an ecotype include "hypothetical" genes with no known function and mobile genetic elements such as phage, transposons, and plasmids, and genes related to transport of ions and metabolites. However, such studies are either based on theoretical ecological niches or have used laboratory-simulated environments to obtain ecotypes within a bacterial community.

The *cpn*60 UT has been established as a target for identification of bacterial species and offers superior resolution at the species and sub-species levels compared to the 16S rRNA gene (Links et al. 2012; Paramel Jayaprakash et al. 2012). In our previous work we showed that *cpn*60 UT sequences permitted the demarcation of phenotypically distinct sub-species groups within *Enterococcus faecalis* and *E. hirae* isolates obtained from pig feces and that this resolution into ecotypes was not observed using 16S rRNA gene sequencing (Vermette et al. 2010). Other studies by our group have shown that *cpn*60 UT sequence relationships predict whole genome sequence similarities between bacteria (Paramel Jayaprakash et al. 2012; Verbeke et al. 2011). However, studies to define genome level differences between ecotypes differentiated by the *cpn*60 UT have not been conducted.

An ecologically motivated genomic comparison should be focused on characterizing gene level differences that could play a role in determining the fitness of a bacterium in a complex microbial community. The objectives of the current study were to identify gene content differences between representative isolates of *cpn*60 defined ecotypes of *E. hirae* from the pig fecal microbial community and to determine if gene content differences are associated with differences in competitiveness of ecotypes. Results of this work provide further support for the application of *cpn*60 UT based methods for high-resolution microbial community profiling and studying bacterial community dynamics.

5.3 Materials and methods

5.3.1 Bacterial strains

Enterococcus hirae isolates listed in Table 5.2 were obtained from feces of healthy pigs as described previously (Vermette et al. 2010). Isolates are named according to pig number-age of pig (weeks)-isolate number (e.g. 88-15-E9). The type strain of *E. hirae* (ATCC 9790) was obtained from the American Type Culture Collection (Manassas, VA). All isolates were grown in tryptone soy broth (TSB) or tryptone soy broth supplemented with 7.5% agar (TSA) at 35ºC.

Table 5.1 Isolates selected for whole genome sequencing and the number of carbon sources utilized by each isolate

*cpn*60 ecotype[1]	Isolate	No. of sole carbon sources utilized
E. hirae-1	78-9-C01	17
	57-9-G06	1
E. hirae-2	81-15-F04	8
	88-15-E09	27
E. hirae-3	57-3-H11	4
	67-3-C05	15

[1](Vermette et al. 2010)

5.3.2 Genomic DNA isolation

Genomic DNA was extracted from isolates using a modified salting out procedure (Martin-Platero et al. 2007), and stored at -20ºC in TE buffer (1 mM EDTA, 10 mM Tris-HCl, pH 7.4). The quality of DNA extracts was assessed by spectrophotometric measurements at 260 and 280 nm, and samples with an A_{260}/A_{280} ratio of 1.8-2.0 were used for analysis. Identity of isolates was confirmed by amplification and sequencing of the *cpn*60 UT region from each DNA extract using primers listed in Table 5.1 and comparison of the resulting sequence to the previously reported sequence for each isolate in cpnDB (www.cpndb.ca) (Vermette et al. 2010).

Table 5.2 List of primers used in the study

Target	Primer	Sequence (5´-3´)	Annealing temperature (° C)
*cpn*60	JH0064	GAT ATY GCW GGW GAY GGW ACA ACD AC	62
	JH0065	CGR CGR TCR CCR AAS CCS GGH GCT	
PTS lac EII CB region 1	JH0483	AACATTCGGTTACTATTTGT	52
	JH0484	TTGGCAATTTAGTATTAAAG	
PTS lac EII CB region 2	JH0485	GATCTTAATGGCTGCTACAGTGA	55
	JH0486	AAGATAGGATTCAATACAAT	
PTS lac EII CB region 3	JH0487	ACGTGCCTCAGCGATTCCCGT	70
	JH0489	CACCCTCAGCTTGATAAGCGGC	
Sodium glutamate symporter	JH0559	ACT GGT TTA TGA CCG ACA	55
	JH0572	CCA ACT CAT TAC CTA GAT AT	
CopB region 1	JH0560	TCC TGG TTC TAA TTG GGT GGT	
	JH0561	ATC TGA TAG AAA CTC TAA	
CopB region 2	JH0562	GTT GAT GAA TCT GCT GTT	
	JH0563	GTT ATC CCC AGT CAA CAT GA	
CopB region 3	JH0564	GGT GTT GGT TTA GAA GCA A	
	JH0565	ATC TTT CAT TTG ATG GAT CAA T	
CopA region 1	JH0566	TTA TTT GAT CGT TTT TCG A	
	JH0567	GGA ATC AGT GGA ACG ATC AAC GGT	
CopA region 2	JH0568	CTG CCT GTT CTA ATT CTA AC	
	JH0569	TCT GCA GCC TTC GCA TTA AGC A	
CopA region 3	JH0570	GCA ATG ATC CGT CCA TCT GTA GGT	
	JH0571	ATG GCA ACG AAT ACC AAA	
Proline racemase	JH0573	TAA TAG TAA GCA GAG TTT TAC T	52
	JH0574	TAT GTG CTG GAT TTT AAT AA	
Cystathione b-lyase	JH0575	GGC ATA ACG GAA ATG AAT GT	60
	JH0576	AAT AAG CCG GAT AAT CCC AA	
D-proline reductase	JH0577	TCC GCG GCG CAT GAT TAT TTC AGA	
	JH0578	TAG GTC TAA CTG ATC TGG CTC TAA	

Target	Primer	Sequence (5´-3´)	Annealing temperature (° C)
Potassium transporting ATPase subunit A	JH0579 JH0580	TAA ATA GAC AAC AAA CAT CTT ATC AAG GAT TAG AAG	52
Potassium transporting ATPase subunit B region 1	JH0581 JH0582	GCT TTC TTT AAT CAT TGT ATT CTT CAG GAT AGG TAT	
Potassium transporting ATPase subunit B region 2	JH0583 JH0584	GTT TAT CTC AAA GAC ATT CAT TGG TAA CTT GAC TAC CAT TAG	
Selenocystine lyase	JH0591 JH0592	GGC CAA AGG AAA AGC ACC CTG TTC GAG TAT CCT TAT CC	55
Selenide water dikinase	JH0593 JH0594	ATA CTT TTC CCA ATG CGT TA TTC TGA TTG TTG GTG TAA C	
Selenocystine tRNA	JH0595 JH0596	TTC AAC AGA ATT AGC TAA AAG ACG ATT ACT GAC TCG A	

5.3.3 Genome sequencing and annotation

Sample preparation and sequencing was done as described in the GS Junior whole genome shotgun sequencing protocol (Roche 454 Life Sciences, Branford, CT), and sequence assembly was done using gsAssembler (v2.7). Genomes were annotated using the Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Resfinder v1.4 (Zankari et al. 2012) was used for identification of acquired anti-microbial resistance genes. CRISPR db (http://crispr.u-psud.fr/crispr/) was used to identify CRISPR sequences. PHAge Search Tool (PHAST) (Zhou et al. 2011) was used to identify phage-like sequences in the genomes of pig fecal isolates.

5.3.4 Comparison of genome sequences

Genome sequences were compared using the Rapid Annotation Using Subsystems Technology (RAST) pipeline (Aziz et al. 2008). JSpecies (v1.2.1), a program that allows for comparison of two or more genomes based on their average nucleotide identity (ANI) and tetra nucleotide frequency (Tetra) values (Richter and Rossello-Mora 2009), was used to determine overall sequence identity between ecotype genomes.

5.3.5 Ecotype specific PCR

Details of the identification of ecotype specific sequences, and design and validation of ecotype specific quantitative PCR assays are presented in the Appendix I. Primers were designed to detect putative ecotype specific sequences using Beacon

Designer (Premier Biosoft, Palo Alto, CA). Specificity of each primer set was checked *in silico* by conducting a BLAST search for each primer against all the genome sequences. The primer sequences, targets, expected product size and annealing temperature for each primer set is given in Table 5.1.

For conventional PCR, each 50 µl reaction mixture contained 1 × PCR reaction buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 2.5 mM $MgCl_2$, 50 pmol of each primer (Table 5.1), 0.2 mM dNTPs (Invitrogen, Burlington, ON), and 2.5 units Taq polymerase (Quanta Biosciences, Gaithersburg, MD). PCR amplifications were done in an Eppendorf Mastercycler EP gradient thermocycler with an initial denaturation step (95ºC for 3 minutes), followed by 40 cycles of denaturation (95ºC, 30 seconds), annealing (temperature in Table 5.1, 30 seconds) and extension (72°C, 1 minute), and a single final extension step (72ºC, 10 minutes). Results of PCR reactions were evaluated by resolution of the products on 1.5% agarose gels stained with ethidium bromide, and visualized and photographed under UV light. Amplified products were purified using EZ-10 spin column PCR purification kit (Bio Basic Inc., Markham, ON) and sequenced using their respective PCR primers (Table 5.1).

5.3.6 Pig fecal extract

Pig fecal extract was prepared by suspending 1000 g of feces from healthy, 9 week old pigs in 2 L of sterile water as described by Hew et al. (2007). The fecal slurry was centrifuged at 24,000 × g for 30 min at 4°C. The supernatant was recovered and centrifuged again at 46,000 × g for 15 min at 4°C. The supernatant was again recovered and filtered using a vacuum driven Millipore GV filter unit (0.22 µm, Millipore,

Billerica, MA). Sterility of the resulting fecal extract was assessed by incubating 2 ml of extract at 37°C with 200 rpm shaking overnight, followed by plating of a 100 µl aliquot of the incubated extract on tryptone soy agar and incubating for an additional 24 hrs at 37°C to confirm sterility. No bacterial colonies were seen on the plates after 24 and 48 hours.

5.3.7 Growth curves

E. hirae isolates were revived on TSA from culture stocks frozen at -80°C. After 24 hours, a loop-full of culture (containing multiple colonies) was transferred to 2 ml of TSB and incubated at 35°C and 200 rpm for 24 hours. For growth curves in fecal extract, fresh fecal extract was inoculated with the 24 hour culture of each isolate such that the final concentration was 10^2 colony forming units per mL (CFU/mL). The inoculated mixture was divided into 2 ml aliquots and incubated at 35°C and 200 rpm. OD_{600} readings were taken at 2, 4, 6, 8, 12 and 24 hrs. Additionally, viable cell culture concentrations were determined by plating of serial dilutions and counting colonies after 24 hours incubation. Growth curves for all strains were done in triplicate. Growth curves were compared by using the *compareGrowthCurves* function from the Statistical Modeling package, statmod, available from the R Project for statistical computing (http://www.r-project.org).

5.3.8 Growth competition

For growth competition assays, each isolate in the competition was co-inoculated into fecal extract to achieve a concentration of 10^2 CFU/mL for each isolate. Inoculated

co-cultures were divided into 2 mL aliquots and incubated at 35°C and 200 rpm. OD_{600} readings were taken at 2, 4, 6, 8, 12 and 24 hrs. A total of 12 competitions were conducted in triplicate. Genomic DNA was extracted from 1 mL of the 24 hour culture and analyzed using ecotype specific qPCR assays to determine the absolute number of cells (represented by copies of assay target gene present) from each ecotype after the competition.

For each competition experiment the mean of log copies/mL was used for the calculation of relative competitive fitness index according to the following formula (Travisano 1997):

Equation 5.1 Competitive fitness index

$$\mathrm{CFI} = \frac{ln\left(\frac{A_f}{A_i}\right)}{ln\left(\frac{B_f}{B_i}\right)}$$

Ai and *Bi* denote the initial quantity of two competing strains; *Af* and *Bf* denote the final quantity of two competing strains after 24 hours. A value of 1 for the relative competitive fitness indicates that strains have grown equally well.

5.4 Results

5.4.1 Selection of isolates for genome sequencing

Two isolates from each ecotype (*E. hirae-1* (78-9-C01 and 57-9-G06), *E. hirae-2* (81-15-F04 and 88-15-E09) and *E. hirae-3* (57-3-H11 and 67-3-C05)) were selected to allow comparison of gene level diversity within and between *E. hirae* ecotypes (Table 5.2). The isolates were selected based on previously reported carbon source utilization

profiles (Vermette et al. 2010). Isolates that utilized a variety of carbon sources within each ecotype were selected to encompass the phenotypic range of the ecotype. This strategy was chosen to allow for as comprehensive a description of the genomic diversity between ecotypes as possible. Additionally, results from our GTG rep PCR genome fingerprinting study (Chapter 3), showed that the minimum percent similarity between genome fingerprints of these six isolates was 50.4 percent (Figure 5.1), further demonstrating genomic diversity among the isolates.

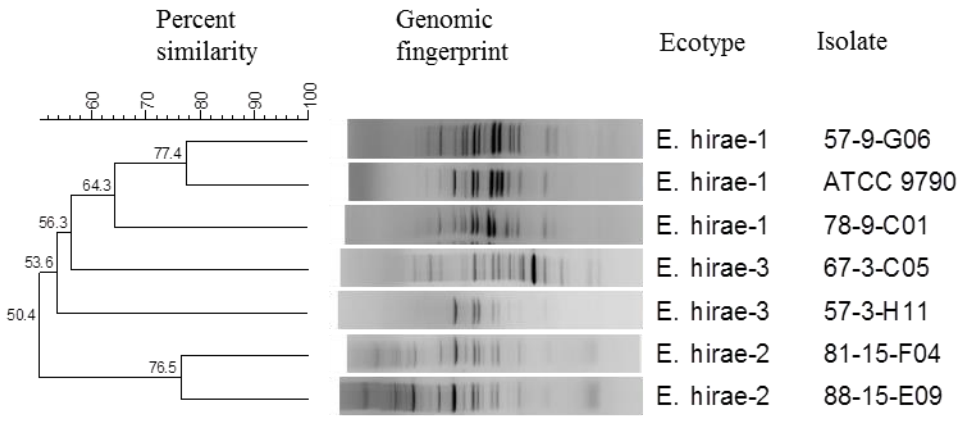


Figure 5.1 Dendrogram of GTG rep PCR based fingerprints of pig fecal *E. hirae* isolates and *E. hirae* type strain (ATCC 9790)

5.4.2 Sequencing results

Whole genome shotgun sequencing of the representative isolates was performed as described in the methods. A comparison of the sequencing runs with instrument benchmarks for whole genome shotgun sequencing is shown in Table 5.3. For all sequencing runs, our results met or exceeded the benchmarks. The number of high quality reads for each genome was >80,000, and percent mixed and dot (defined as more than one read or no read per bead) was ≤20%. Similarly, the percentages of reads with too short quality and too short primer (less than the minimum length defined) were <30% and <1% respectively.

Table 5.3 Comparison of genome sequencing run results with Roche GS Junior benchmarks

cpn60 defined ecotype		***E. hirae-1***		***E. hirae-2***		***E. hirae-3***	
Parameters	**Benchmark**	**57-9-G06**	**78-9-C01**	**81-15-F04**	**88-15-E09**	**57-3-H11**	**67-3-C05**
Raw wells	≤250,000	244,842	196,867	150,050	242,424	231,658	112,841
Keypass Wells	>90%	98.8	95.8	95.0	97.3	97.9	93.9
HQ reads	50,000-80,000	103,063	147,873	120,872	108,749	99,020	90,959
% filter passed	>50%	42.6	78.0	84.8	46.0	43.7	85.9
% Mixed + dots	<20%	19.9	3.6	4.9	19.6	20.2	5.4
% TSQ[1]	<30%	35.7	16.7	10.3	31.1	36.1	8.7
% TSP[2]	<2%	<1	<1	<1	<1	<1	<1

[1] TSQ=Too short quality
[2] TSP=Too short primer

5.4.3 General genome features

General features of the draft genomes of all the pig fecal isolates in comparison to the *E. hirae* type strain ATCC 9097 genome sequence are presented in Table 5.4. The sequences have been deposited in Genbank under BioProjects PRJNA172133, PRJNA182154, PRJNA182156, PRJNA182157, PRJNA175366 and PRJNA175367. Coverage estimates were obtained by dividing the number of sequenced and assembled bases with the estimated *E. hirae* genome size of 2.8 Mb (size of *E. hirae* ATCC 9790 type strain genome). The coverage for all isolates ranged from 15× to 24×, corresponding to a "standard draft" quality genome (Chain et al. 2009). The %G+C content for all pig fecal isolates was 36.3 to 37.1%, similar to the *E. hirae* type strain (36.9%) and *Enterococcus* as a genus (37 to 45%) (Lebreton et al. 2014). Average gene length, number of protein coding genes and percentage of the ORFs with a functional assignment were also similar across all six pig fecal isolate genomes and with the *E. hirae* type strain (Table 5.4). Only one complete sequence was observed for each of the rRNA genes: 16S rRNA, 23S rRNA and 5S rRNA. This is contrary to the observation that in the finished genome of the *E. hirae* type strain there are six copies of the ribosomal RNA operon. The discrepancy is likely due to the nature of the shotgun sequencing and assembly process, which automatically collapses identical sequence regions, which would artificially reduce the number of identical rRNA genes to one.

Table 5.4 Genome features of pig fecal *E. hirae* isolates and *E. hirae* type strain[1]

***cpn*60 defined ecotype**	***E. hirae-1***		***E. hirae-2***		***E. hirae-3***		**Type strain (*E. hirae-1*)**
Genome features	**57-9-G6**	**78-9-C1**	**81-15-F4**	**88-15-E9**	**57-3-H11**	**67-3-C5**	**ATCC9790**
Size (bp)	2914949	2928799	2852997	2860996	2793003	2849590	2856440
No. of contigs (>500 bp)	98	62	76	72	102	149	1
N50	61682	112644	67326	70870	77040	49718	
%GC	36.5	37.1	38.4	36.3	36.3	36.9	36.9
Sequencing coverage (×)	15	24	19	17	15	15	Complete
Protein-coding genes							
Total	2743	2819	2623	2566	2601	2677	2756
No. (%) with functional assignment	1721 (62.7)	1825 (64.7)	1669 (63.6)	1823 (71.0)	1694 (65.1)	1727 (64.5)	2064 (72.5)
No. (%) hypothetical	1022 (37.2)	994 (35.2)	954 (36.3)	743 (28.9)	907 (34.8)	950 (35.4)	692 (24.3)
RNA							
No. of rRNA genes[2]	3	3	3	3	3	3	18
16S rRNA	1	1	1	1	1	1	6*
No. of tRNA genes	53	57	47	54	51	53	71

[1] *E. hirae* type strain (ATCC 9790) is included as a reference

[2] 16S rRNA, 23S rRNA and 5S rRNA genes

*The discrepancy is likely due to the nature of the shotgun sequencing and assembly process, which automatically collapses identical sequence regions, which would artificially reduce the number of identical rRNA genes to one

Genome sizes, estimated as the number of assembled bases, were 2.8 to 2.9 Mb. For all isolates except 57-3-H11 (*E. hirae-3*), this estimated genome size was larger than the *E. hirae* type strain chromosome. However, the type strain is known to harbor a 28,699 bp plasmid. Thus, the variation in size between pig fecal *E. hirae* strains and the type strain could be due to presence of plasmid sequences in addition to other unique DNA sequences in the genomes. Plasmid-associated replication initiation genes *repA*, *repB*, and *repC* were annotated in the genomes of all the pig fecal isolates, suggesting that plasmids are present.

To obtain the genome similarities between and within the sequenced *E. hirae* isolates, tetranucleotide score (Tetra), average nucleotide identity by BLAST (ANIb) and average nucleotide identity by MUMMer (ANIm) were calculated for all *E. hirae* isolates, *E. faecalis* strain V583 and the *E. hirae* type strain using JSpecies (v 1.2.1) (Table 5.5). All values were consistent with cutoffs proposed by Richter and Rossello-Mora (2009) as indicative of inter- and intra-specific genome similarity (>95-96% ANIm and ANIb, >99% Tetra indicate same species).

Table 5.5 Tetra (first row), ANIb (second row) and ANIm (third row) scores for *E. hirae* isolates, *E. faecalis* V583 and *E. hirae* type strain (ATCC 9790)[1]

***cpn*60 ecotype**	***E. hirae-1***	***E. hirae-2***		***E. hirae-3***		***E. hirae-1***	***E. faecalis***
Isolate ID	78-9-C01	81-15-F04	88-15-E09	57-3-H11	67-3-C05	ATCC 9790	V583
57-9-G6	0.99991	0.99951	0.99953	0.99939	0.99941	0.99942	0.91146
	99.94	98.76	98.76	98.45	98.46	99.7	71.13
	99.88	98.8	98.81	98.81	98.83	99.75	86.02
78-9-C1		0.99953	0.99952	0.99945	0.99946	0.99955	0.91206
		98.79	98.78	98.47	98.46	99.72	71.3
		98.82	98.81	98.81	98.83	99.77	85.95
81-15-F4			0.99997	0.99925	0.99934	0.99964	0.90975
			99.84	98.74	98.65	98.75	71.3
			99.93	98.93	98.89	98.84	85.54
88-15-E9				0.99927	0.99937	0.99962	0.9099
				98.87	98.73	98.83	71.13
				98.93	98.88	98.81	85.47
57-3-H11					0.99954	0.99916	0.90776
					98.94	98.7	71.35
					99	98.86	87.04
67-3-C5						0.99923	0.9125
						98.49	72.25
						98.88	88.08
***E. faecalis* V583**						0.90998	
						71.39	
						86.99	

[1] Grey shading indicates values consistent with intra-specific comparisons.

Clusters of Orthologous Groups (COGs) are categories of proteins related by function (Tatusov et al. 2001). To compare the distribution of functional genes, predicted proteomes of isolates within an ecotype were pooled and their COG representation based on PGAAP annotations were compared. The genomes of pig fecal isolates and the type strain had nearly identical predicted functional repertoires, with the largest variation in number of COGs in functional category L (replication, recombination and repair) (Figure 5.2). The maximum number of unique COGs for each *E. hirae* ecotype belonged to category S (function unknown), followed by categories R (general function prediction) and G (carbon uptake and metabolism) (Figure 5.2).

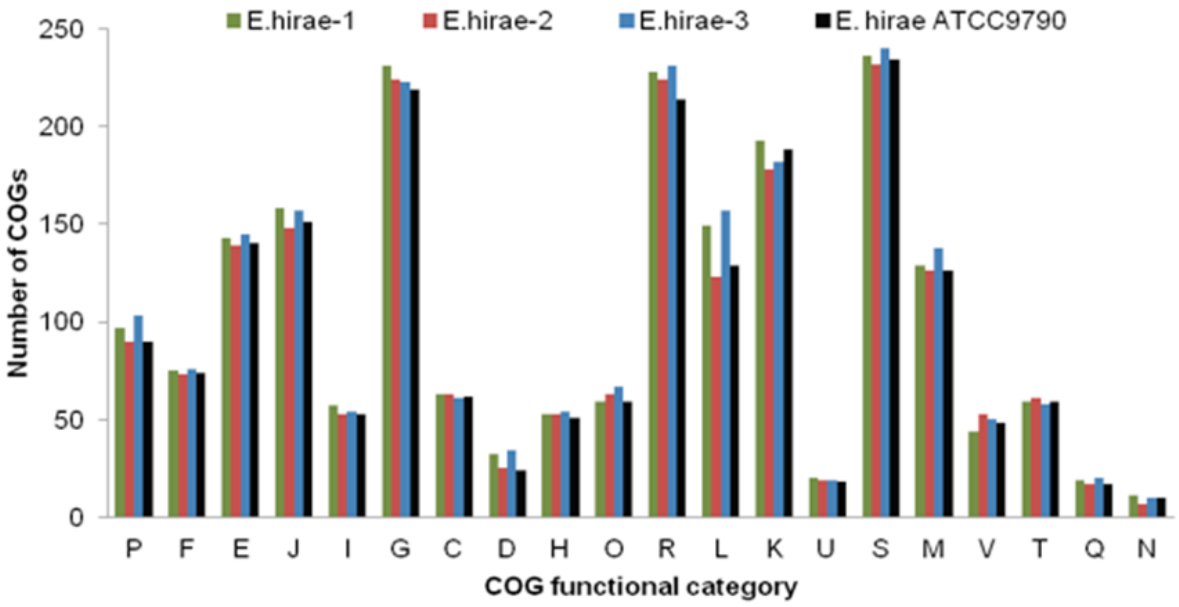
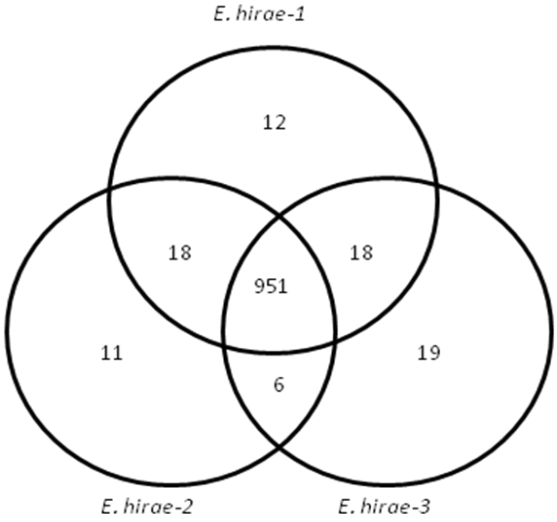


Figure 5.2 Distribution of COGs in all pig fecal *E. hirae* ecotypes and *E. hirae* type strain ATCC 9790.

Isolates from the same ecotype were combined and reduced to a unique COG list for this analysis. COG categories: J: translation, ribosomal structure and biogenesis; K: transcription; L: replication, recombination and repair; D: cell cycle control, cell division, chromosome partitioning; V: defense mechanisms; T: signal transduction mechanisms; M: cell wall/membrane/envelope biogenesis; N: cell motility; U: intracellular trafficking, secretion, and vesicular transport; O: posttranslational modification, protein turnover; chaperones; P: inorganic ion transport and metabolism; C: energy production and conversion; G: carbohydrate transport and metabolism; E: amino acid transport and metabolism; F: nucleotide transport and metabolism; H: coenzyme transport and metabolism; I: lipid transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown.

A total of 951 COGs were shared between the sequenced *E. hirae* isolates and only 12, 11 and 19 COGs were uniquely identified in isolates from *E. hirae*-1, *E. hirae-2* and *E. hirae-3* respectively (Figure 5.3a). A pair-wise comparison of COGs from all pig fecal *E. hirae* isolates and the type strain showed that 944 COGs were shared by all isolates whereas 14 and 7 COGs were unique to the type strain and the pig fecal *E. hirae* strains, respectively (Figure 5.3b). A list of unique COGs for each ecotype is presented in Table 5.6. Unique COGs from category G (carbon uptake and metabolism) were detected in all ecotypes (Table 5.6), suggesting that proteins involved in carbon uptake and metabolism play an important role in genotypic and phenotypic diversity in *E. hirae* ecotypes.

A



B

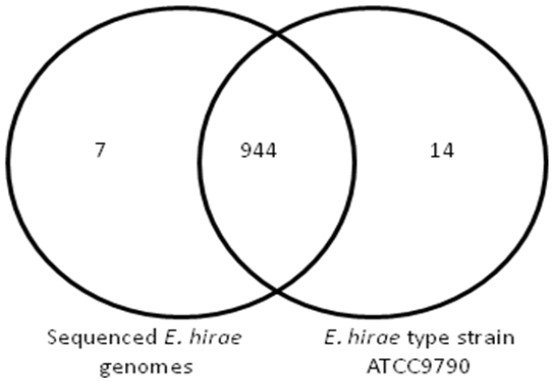


Figure 5.3 Distribution of shared and unique COGs
(A) among pig fecal *E. hirae* strains, and (B) between all six pig fecal *E. hirae* isolates and the *E. hirae* type strain. Isolates from the same ecotype were combined and reduced to a unique COG list for this analysis.

Table 5.6 Unique COGs in *E. hirae* ecotypes

***cpn*60 defined ecotype**	**COG**	**Category[1]**	**Function**
E. hirae-1	COG0425	O	Predicted redox protein, regulator of disulfide bond formation
	COG0433	R	Predicted ATPase
	COG0840	N	Methyl-accepting chemotaxis protein
	COG1112	L	Superfamily I DNA and RNA helicases and helicase subunits
	COG2971	G	Predicted N-acetylglucosamine kinase
	COG3276	J	Selenocysteine-specific translation elongation factor
	COG3291	R	PKD repeat proteins
	COG4043	S	Uncharacterized protein conserved in bacteria
	COG4115	S	Uncharacterized protein conserved in bacteria
	COG4640	S	Predicted membrane protein
	COG5412	S	Phage-related protein
	COG5655	L	plasmid rolling circle replication initiator protein and truncated derivatives
E. hirae-2	COG1715	V	Restriction endonuclease
	COG1783	R	Phage terminase large subunit
	COG1932	HE	Phosphoserine aminotransferase
	COG2902	E	NAD-specific glutamate dehydrogenase
	COG3177	S	Uncharacterized conserved protein
	COG3386	G	Gluconolactonase
	COG3561	K	Phage anti-repressor protein
	COG3600	S	Uncharacterized phage-associated protein
	COG4570	L	Holliday junction resolvase
	COG4653	R	Predicted phage phi-C31 gp36 major capsid-like protein
	COG4707	S	Uncharacterized protein conserved in bacteria
E. hirae-3	COG0071	O	Molecular chaperone (small heat shock protein)
	COG0378	OK	Ni2+-binding GTPase involved in regulation of expression and maturation of urease and hydrogenase

*cpn*60 defined ecotype	COG	Category[1]	Function
	COG0390	R	ABC-type uncharacterized transport system, permease component
	COG0615	MI	Cytidylyltransferase
	COG0662	G	Mannose-6-phosphate isomerase
	COG1061	KL	DNA or RNA helicases of superfamily II
	COG1404	O	Subtilisin-like serine proteases
	COG1555	L	DNA uptake protein and related DNA-binding proteins
	COG1887	M	Putative glycosyl/glycerophosphate transferases involved in teichoic acid biosynthesis TagF/TagB/EpsJ/RodC
	COG2184	D	Protein involved in cell division
	COG3588	G	Fructose-1,6-bisphosphate aldolase
	COG3774	M	Mannosyltransferase OCH1 and related enzymes
	COG4235	O	Cytochrome c biogenesis factor
	COG4619	R	ABC-type uncharacterized transport system, ATPase component
	COG4983	S	Uncharacterized conserved protein
	COG5017	S	Uncharacterized conserved protein
	COG5340	K	Predicted transcriptional regulator
	COG5492	N	Bacterial surface proteins containing Ig-like domains
	COG5519	L	Superfamily II helicase and inactivated derivatives

Isolates from the same ecotype were combined and reduced to a unique COG list for this analysis.

[1]COG categories: J: translation, ribosomal structure and biogenesis; K: transcription; L: replication, recombination and repair; D: cell cycle control, cell division, chromosome partitioning; V: defense mechanisms; T: signal transduction mechanisms; M: cell wall/membrane/envelope biogenesis; N: cell motility; U: intracellular trafficking, secretion, and vesicular transport; O: posttranslational modification, protein turnover; chaperones; P: inorganic ion transport and metabolism; C: energy production and conversion; G: carbohydrate transport and metabolism; E: amino acid transport and metabolism; F: nucleotide transport and metabolism; H: coenzyme transport and metabolism; I: lipid transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown.

5.4.4 Lactose utilization and phosphoenolpyruvate (PEP)-dependent phosphotransferase system

Results from our previous study showed that 92% (36/39) of *E. hirae-2* isolates could utilize lactose as a sole carbon source in comparison to only 16% (10/64) of *E. hirae-1* isolates and none (0/5) of the *E. hirae-3* isolates (Vermette et al. 2010). To explain these differences in lactose utilization, and to follow up on the observation of COG category G differences between ecotypes, we compared the putative genes involved in lactose uptake and metabolism in all the pig fecal *E. hirae* isolates. Lactose metabolism has been studied extensively in both Gram positive and Gram negative bacteria (de Vos and Vaughan 1994; Ozbudak et al. 2004), and occurs *via* two different pathways (Kandler 1983). The first pathway involves a carbohydrate-specific permease (lac Y) that imports lactose into the intracellular space for hydrolysis by β-galactosidase into galactose and glucose. The second pathway of lactose utilization occurs through lactose-specific PTS transporters, which are multi-component systems used by bacteria for uptake of carbohydrates. Each PTS consists of non-specific components: enzyme I (EI) and histidine phosphocarrier protein, and a number of sugar specific permeases (enzyme II). Each enzyme II (EII) complex consists of one or two hydrophobic integral membrane domains (domains C and D) and two hydrophilic domains (domains A and B). EII complexes may exist as separate proteins or as a single, multi-domain protein. To date, little information is available about lactose uptake in the genus *Enterococcus*, although the involvement of PTS has been suggested (Kandler 1983).

A putative lactose permease (lac Y) gene was not annotated in any of the *E. hirae* genomes. To further confirm this observation, the sequences of lac Y genes from *Escherichia coli* strain K12 MG1655 (NC_000913.3) and *Bacillus megaterium* (NC_014019.1) were queried against genome sequences of all the *E. hirae* ecotypes using BLASTn. No sequences with significant matches were obtained suggesting that these *E. hirae* isolates lack a lac Y homolog. Based on this information, we focused on studying diversity in the lac family PTS among *E. hirae* isolates.

As expected, only one gene for each of EI and HPr (the non-specific/general components of PTS) was detected in the genome sequences of the six *E. hirae* isolates. According to the Transporter Classification Database (www.tcdb.org) and phylogenetic analysis of EII components, the EII permeases can be classified into seven families: glucose (including glucoside) (Glc), fructose (including mannitol) (Fru), lactose (including N,N-diacetylchitobiose) (Lac), galactitol (Gat), glucitol (Gut), mannose (Man), and L-ascorbate (Asc). The percentage of total predicted open reading frames corresponding to putative PTS component sequences ranged from 2.2 to 2.3% between the pig fecal isolates, and PTS EII components for all seven EII families were present in isolates (Table 5.7). The order of abundance of EII components was Glc > Lac > Man > Fru > Gut > Gat > Asc. PTS EII protein domain fusions were identified based on the PGAAP annotation and are listed in Table 5.8. Fusion domains EI-HPr, EI-EII (A or B or C) components and HPr-EII (A or B or C) components were not found in any of the pig fecal *E. hirae* genomes.

Table 5.7 Number of PTS EII components by family in pig fecal *E. hirae* isolates

***cpn* 60 defined ecotype**	**Isolate**	**PTS EII family**							**Total EII components**	**Percent of total ORFs**
		Glc	Fru	Lac	Gut	Gat	Man	Asc		
E. hirae-1	57-9-G6	16	11	17	3	4	12	1	64	2.4
	78-9-C1	14	13	13	6	1	13	1	61	2.2
E. hirae-2	81-15-F4	16	14	19	3	4	9	1	66	2.6
	88-15-E9	16	7	17	3	4	13	1	61	2.5
E. hirae-3	57-3-H11	22	12	10	6	3	15	1	69	2.7
	67-3-C5	17	11	15	3	3	14	1	64	2.5
	Total EII components	101	68	91	24	19	76	6	385	

Table 5.8 Domain fusions for EII complexes observed in pig fecal *E. hirae* isolates

Domain fusion	PTS Enzyme II complex family	Predicted carbon source(s) transported
AB	Man	mannose/fructose/N-acetylgalactosamine
ABC	Glc	sucrose
ABC	Glc	β-glucosides
ABC	Fru	glucose/maltose
ABC	Fru	mannitol/fructose
ABC	Fru	fructose
ABC	Glc	glucose/maltose/N-acetylglucosamine
ABC	Glc	glucose/maltose
BC	Glc	trehalose
BC	Fru	fructose
BC	Gut	glucitol/sorbitol
BC	Glc	sucrose
BC	Fru	mannitol/fructose
BC	Gut	sorbitol
BC	Fru	mannitol/fructose
BCA	Glc	sucrose
BCA	Glc	glucose/maltose
BCA	Glc	glucose/maltose/N-acetylglucosamine
BCA	Glc	maltose
CB	Lac	cellobiose/lactose
CB	Glc	glucose/maltose/N-acetylglucosamine

To investigate the difference in lactose utilization between *E. hirae* ecotypes, we focused on studying the diversity in lac family EII PTS components. The numbers of lac family EII components (A, B and C) in *E. hirae* isolates are shown in Table 5.9. It has been suggested that EII C components are highly specific for the carbon source they transport while EII A and EII B components can be shared between several carbon sources, which explains the relatively low number of EII A (range 3-6 per genome) and EII B (4-5) components in comparison to EII C components (7-10) detected. Similar observations have been made by Cases et al. (2007) for Firmicutes in general.

Table 5.9 Number of ORFs predicted to encode EII components for Lac family PTS observed in pig fecal *E. hirae* isolates

	E. hirae-1		*E. hirae-2*		*E. hirae-3*	
EII component	**57-9-G6**	**78-9-C1**	**81-15-F4**	**88-15-E9**	**67-3-C5**	**57-3-H11**
A	4	4	6	4	3	4
B	4	5	5	5	5	4
C	9	9	10	9	8	7

To understand the distribution of lac family PTS EII components among *E. hirae* ecotypes, a phylogenetic analysis was conducted. None of the lac family PTS EII components observed in *E. hirae* pig isolate genomes or the *E. hirae* type strain ATCC 9790 were annotated as lactose specific, so all EII lac family components were considered putatively lactose specific and included in the analysis. Phylogenetic trees of predicted amino acid sequences of each EII components (A, B and C) are shown in Figures 5.4-5.6, respectively. The sequences formed distinct clusters containing orthologous sequences from multiple isolates. Since EII A and EII B components can be shared for transport of multiple lac family carbon sources (lactose, cellobiose and N,N'-diacetylchitobiose and other aromatic β-glucosides), we further investigated lac family EII C components to identify gene content differences that might explain the different lactose utilization phenotypes among ecotypes.

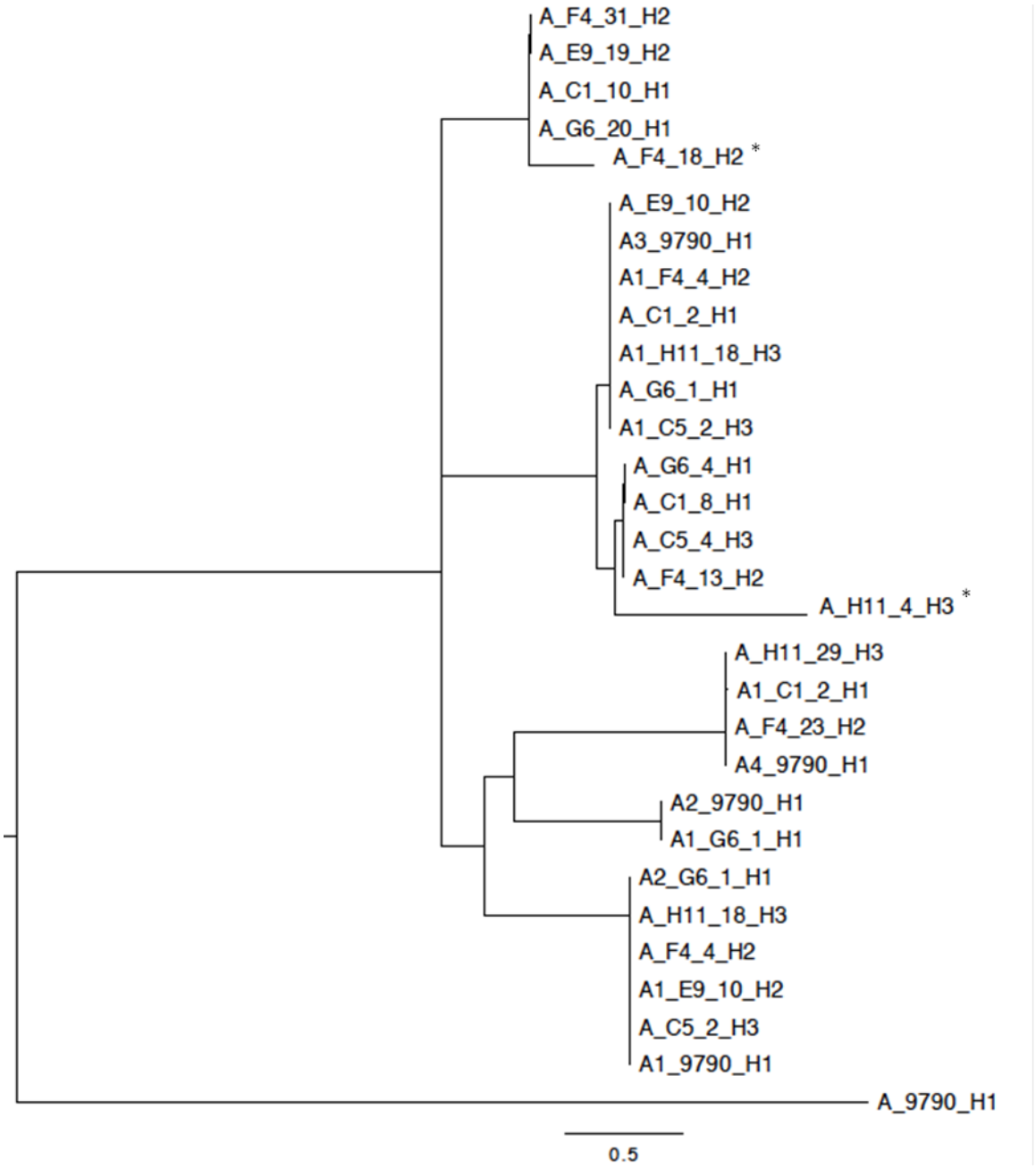


Figure 5.4 Phylogenetic tree of 30 putative lac family EII A subunits of pig fecal *E. hirae* isolates and *E. hirae* type strain ATCC 9790

All nodes have 100% bootstrap support. Asterisks indicate sequences containing predicted frameshifts. Each branch in the phylogenetic tree represents an amino acid sequence from an isolate and is named as EII component_isolate ID_contig number_ecotype.

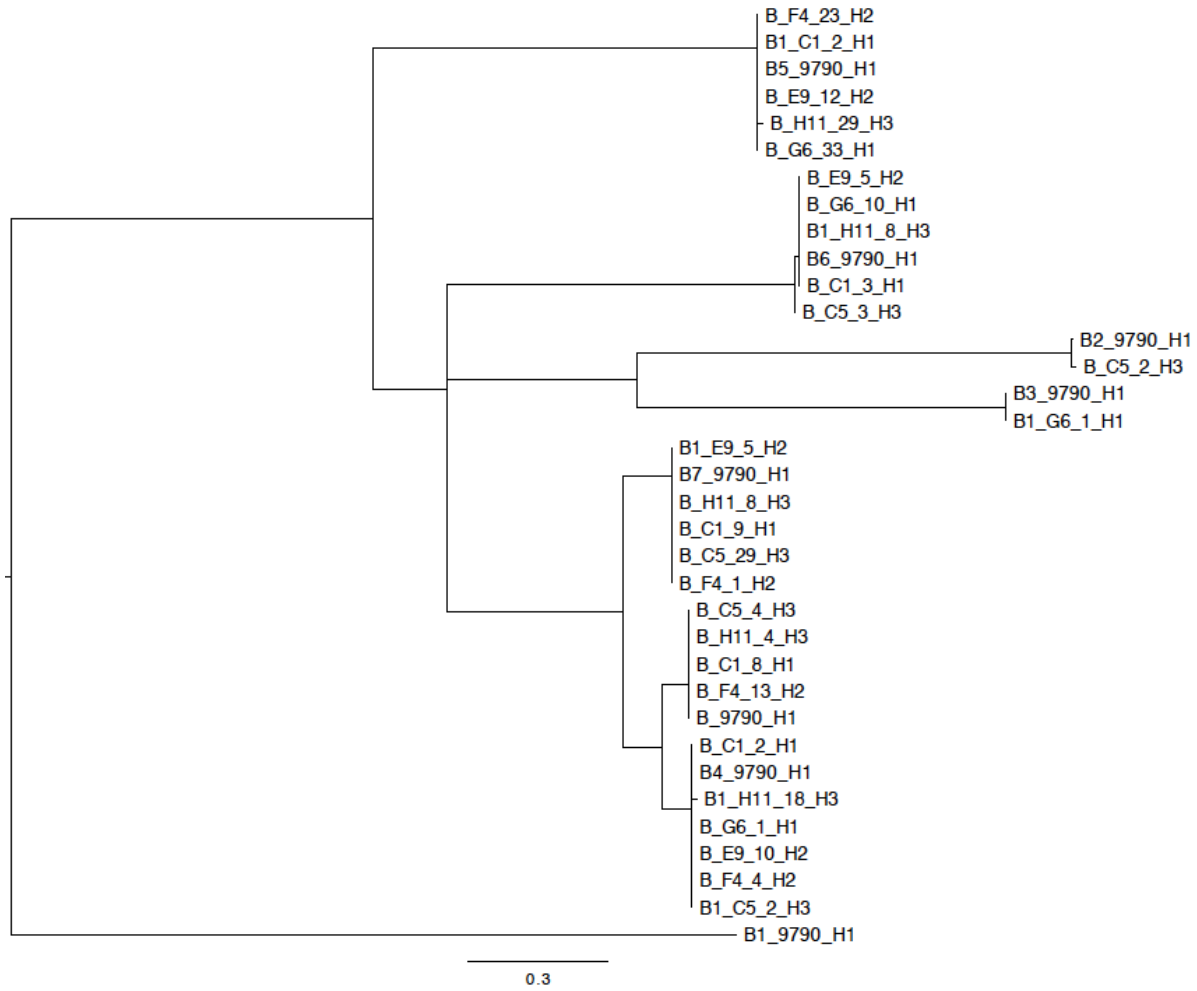


Figure 5.5 Phylogenetic tree of 35 putative lac family EII B subunits of pig fecal *E. hirae* isolates and the *E. hirae* type strain ATCC 9790

All nodes have 100% bootstrap support. Each branch in the phylogenetic trees represents an amino acid sequence from an isolate and is named as EII component_isolate ID_contig number_ecotype.

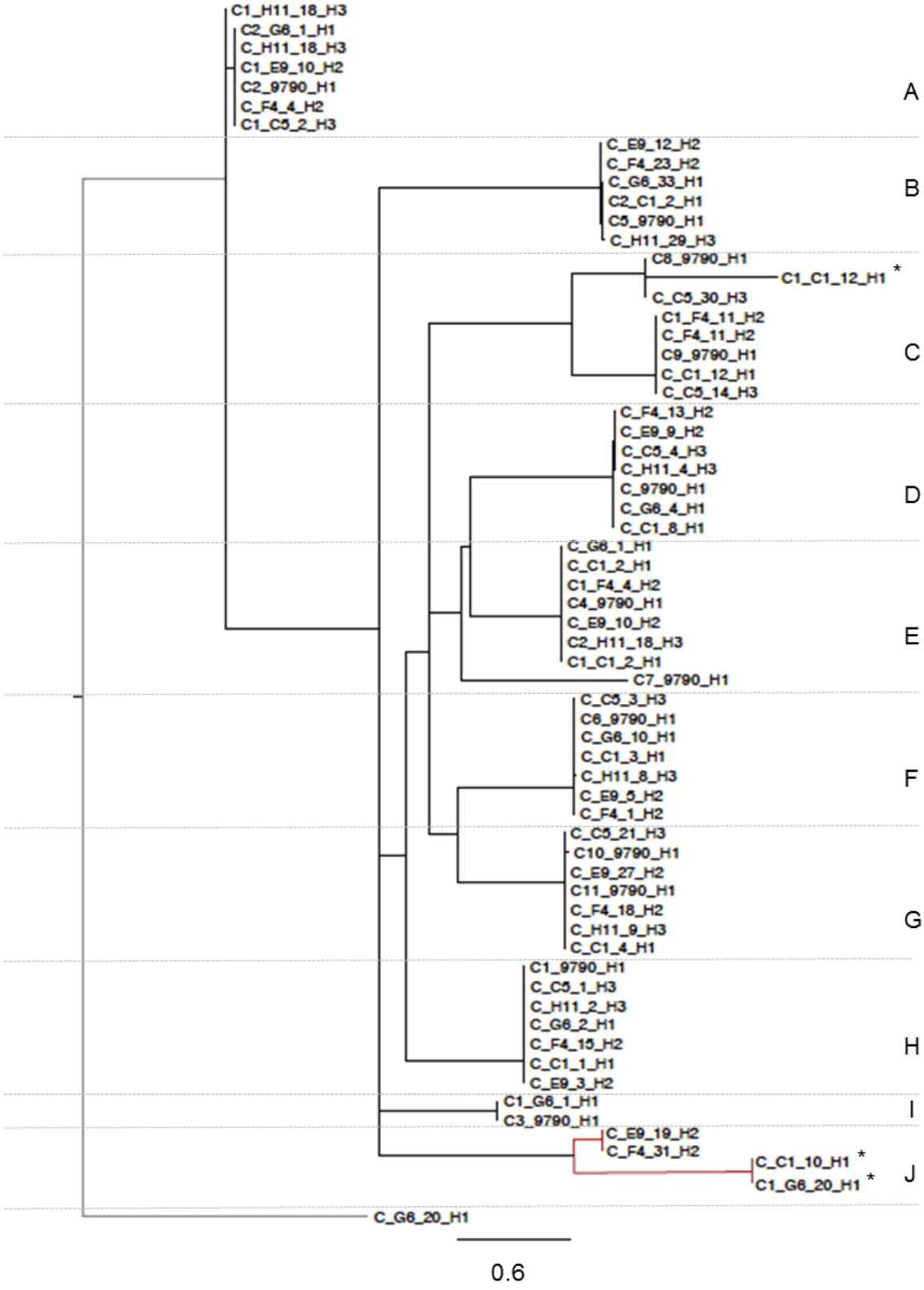

C1_H11_18_H3
C2_G6_1_H1
C_H11_18_H3
C1_E9_10_H2
C2_9790_H1
C_F4_4_H2
C1_C5_2_H3
A
C_E9_12_H2
C_F4_23_H2
C_G6_33_H1
C2_C1_2_H1
C5_9790_H1
C_H11_29_H3
B
C8_9790_H1
C1_C1_12_H1 *
C_C5_30_H3
C1_F4_11_H2
C_F4_11_H2
C9_9790_H1
C_C1_12_H1
C_C5_14_H3
C
C_F4_13_H2
C_E9_9_H2
C_C5_4_H3
C_H11_4_H3
C_9790_H1
C_G6_4_H1
C_C1_8_H1
D
C_G6_1_H1
C_C1_2_H1
C1_F4_4_H2
C4_9790_H1
C_E9_10_H2
C2_H11_18_H3
C1_C1_2_H1
C7_9790_H1
E
C_C5_3_H3
C6_9790_H1
C_G6_10_H1
C_C1_3_H1
C_H11_8_H3
C_E9_5_H2
C_F4_1_H2
F
C_C5_21_H3
C10_9790_H1
C_E9_27_H2
C11_9790_H1
C_F4_18_H2
C_H11_9_H3
C_C1_4_H1
G
C1_9790_H1
C_C5_1_H3
C_H11_2_H3
C_G6_2_H1
C_F4_15_H2
C_C1_1_H1
C_E9_3_H2
H
C1_G6_1_H1
C3_9790_H1
I
C_E9_19_H2
C_F4_31_H2
C_C1_10_H1 *
C1_G6_20_H1 *
J
C_G6_20_H1
0.6

Figure 5.6 Phylogenetic tree of 64 putative lac family EII C subunits of pig fecal *E. hirae* isolates and *E. hirae* type strain ATCC 9790.
All nodes have 100% bootstrap support. Asterisks indicate sequences containing predicted frameshifts. Each branch in the phylogenetic tree represents an amino acid sequence from an isolate and is named as EII component_isolate ID_contig number_ecotype. Letters represent clusters of identical EII C sequences. The branches in red indicate the lactose specific EII CB sequences (cluster J).

All of the clusters of lac family EII C components except clusters I and J contain sequences from each pig fecal isolate (Figure 5.6). Sequences in cluster J corresponded to an EII CB fusion domain detected only in *E. hirae- 1* and *E. hirae-2* isolates, but not in *E. hirae-3*. Based on the sequence alignment, this EII CB fusion domain was apparently disrupted by a frameshift in *E. hirae-1* isolates. EII CB specific PCR using primers JH0483-JH0489 (Table 5.1) was used to confirm the distribution of cluster J EII CB sequences among the fecal isolate genomes. PCR products of the expected size were observed for isolates in *E. hirae-1* and *E. hirae-2* ecotypes, but no amplification was observed from *E. hirae-3* isolates, confirming the absence of the EII CB fusion domain in this ecotype. The purified PCR products were sequenced and results confirmed the frameshift in EII CB *E. hirae-1* isolates. Examination of the genomic context of the putative lactose specific EII CB domain in *E. hirae-2* suggested that the gene occurs in an operon with an EII A component, a 6-phospho-β-galactosidase and a lactose regulator (Figure 5.7). The operon was disrupted in *E. hirae-1* genomes.

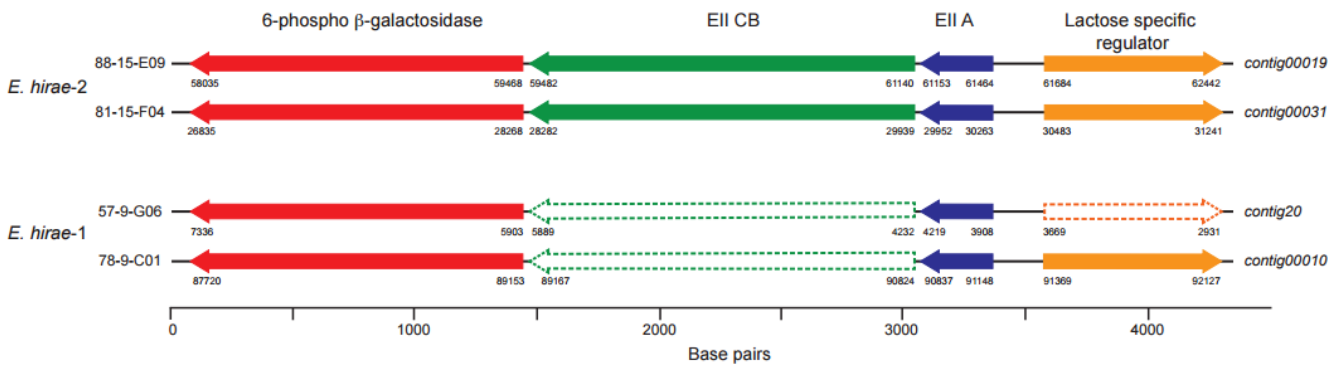


Figure 5.7 Genomic region containing the putative PTS EII CB fusion domain gene in *E. hirae*-1 and *E. hirae*-2 genomes.
Nucleotide numbers at the start and end of each ORF are relative to the indicated contig number in each assembled genome. Arrows outlined with broken lines are affected by frame-shifts (EII CB fusion domain) or in-frame stop codon (lactose specific regulator). This operon was not detected in *E. hirae-3*.

To determine the broader distribution of the putative lactose specific EII CB fusion domains, we screened all the *E. hirae* isolates in our culture collection (*E. hirae-1* (n=61), *E. hirae-2* (n=39), *E. hirae-3* (n=5) using EII CB domain specific PCR. PCR products of the expected size were seen for all *E. hirae-1* isolates (61/61), 37/39 *E. hirae-2* isolates, and none of the five *E. hirae-3* isolates. PCR products from a total of 20 *E. hirae-1* (n=14) and *E. hirae-2* (n=6) isolates were selected based on utilization of lactose as the sole carbon source and sequenced. All of the *E. hirae-2* isolates were lactose positive and all had uninterrupted EII CB fusion domain gene sequences. The expected frameshift was observed in all *E. hirae-1* isolates, although 6/14 had been previously reported to utilize lactose as a sole carbon source (Vermette et al. 2010).

5.4.5 Other metabolic pathways

5.4.5.1 Amino acid metabolism

i) Glutamine, glutamate, aspartate and asparagine metabolism - Sodium/glutamate symporter

The genomes of all four isolates in *E. hirae-1* and *E. hirae-2* ecotypes have a 1377 bp ORF that encodes a putative Na^+/glutamate symporter but only a part of this sequence was detected in *E. hirae-3* isolates (145/1377 and 745/1377 nt are present in *E. hirae*-3 isolates 57-03-H11 and 67-03-C5, respectively), suggesting a deletion in the ORF for putative Na^+/glutamate symporter. ORFs encoding a putative translation initiation factor, a putative coenzyme A disulfide reductase and a conserved hypothetical protein that are present adjacent to the Na^+/glutamate symporter in *E. hirae -1* and *E. hirae-2* are also missing from the genomes of both the *E. hirae-3* isolates. These results suggest that

E. hirae-3 isolates cannot transport L-glutamate into the cytosol *via* a Na^+/glutamate symporter. Interestingly, genes putatively encoding enzymes involved in glutamate, aspartate, glutamine and asparagine metabolism (inferred from BioCyc pathways): glutamine synthetase type I (EC 6.3.1.2), NADP-specific glutamate dehydrogenase (EC 1.4.1.4), glutamate racemase (EC 5.1.1.3) and glutaminase (EC 3.5.1.2), asparagine synthetase (EC 6.3.5.4), aspartate aminotransferase (EC 2.6.1.1), aspartate racemase (EC 5.1.1.13), L-asparaginase (EC 3.5.1.1), were detected in all the genomes. We confirmed ecotype specific distribution of the putative Na^+/glutamate symporter gene by PCR as explained in the methods section. PCR products of the expected size were observed in 12/17 *E. hirae-1*, 7/17 *E. hirae-2* and 0/5 of the *E. hirae-3* isolates. The PCR products from two isolates of *E. hirae-1* and *E. hirae-2* were sequenced to confirm the identity of the PCR products.

ii) *Proline metabolism*

A complete set of genes encoding proteins required for synthesis of proline was detected in all *E. hirae-1* and *E. hirae-2* isolates: Proline dehydrogenase (EC 1.5.99.8), pyrroline-5-carboxylate reductase (EC 1.5.1.2), gamma-glutamyl phosphate reductase (EC 1.2.1.41), glutamate 5-kinase (EC 2.7.2.11), pyrroline-5-carboxylate reductase (EC 1.5.1.2), D-proline reductase (EC 1.21.4.1) and proline racemase (EC 5.1.1.4). However, putative D-proline reductase (EC 1.21.4.1) and proline racemase (EC 5.1.1.4) were not detected in *E. hirae-3* isolates, suggesting either the inability of these isolates to synthesize proline or presence of an alternate pathway. We confirmed this ecotype specific distribution of the putative D-proline reductase and putative proline racemase genes by PCR. For the putative D-proline reductase gene, PCR products of the expected

size were observed for 9/17 *E. hirae-1*, 11/17 *E. hirae-2* and 0/5 *E. hirae-3* isolates. Similarly, for the putative proline racemase gene, a PCR product of expected size was observed in 14/17 *E. hirae-1*, 17/17 *E. hirae-2* and none of the *E. hirae-3* (n=5) isolates. Positive PCR products from two isolates of *E. hirae-1* and *E. hirae-2* were sequenced to confirm the identity of the PCR products.

iii) *Methionine metabolism*

Methionine synthesis involves conversion of L-cystathionine to L-homocysteine and finally to L-methionine. In *E. hirae* ATCC 9790 the annotated cystathionine beta-lyase gene (EHR_09290) is 1140 bp in length (encoding 370 aa). A homologous sequence was annotated only in *E. hirae-1* and *E. hirae-2* isolates, suggesting that the pathway for methionine biosynthesis is incomplete in *E. hirae-3* isolates. Ecotype specific distribution of the putative cystathionine beta-lyase gene was confirmed by PCR. A PCR product of the expected size was observed for 16/17 *E. hirae-1*, 15/17 *E. hirae-2* and none of the *E. hirae-3* (n=5) isolates. Positive PCR products from two isolates of *E. hirae-1* and *E. hirae-2* were sequenced to confirm the identity of the PCR products.

iv) *Selenocysteine metabolism*

A putative operon including genes for selenocysteine biosynthesis comprising selenide water dikinase (EC 2.7.9.3), seryl-t-RNA synthetase (EC 6.1.1.11) and selenocysteine synthase (EC 2.9.1.1) along with the selenocysteine specific insertion sequence and an elongation factor were detected only in *E. hirae-1* and *E. hirae* ATCC 9790 (affiliated with *E. hirae-1* based on *cpn*60 UT sequence). PCR targeting the putative selenide water dikinase, seryl-t-RNA synthetase and selenocysteine synthase genes confirmed this ecotype-specific distribution. All three genes were detected in 14/17 *E.*

hirae-1 isolates and none of the *E. hirae-2* (n=17) or *E. hirae-3* (n=5) isolates. Positive PCR products from two isolates of *E. hirae-1* were sequenced to confirm the identity of the PCR products.

5.4.5.2 Potassium-transporting ATPases

Potassium is one of the most abundant intracellular cations. It is essential for regulation of cytosolic pH and is a cofactor for several enzymes. In *E. hirae*, potassium uptake involves two potassium transporting genes: KtrI (EHR_00050) and KtrII (EHR_03990) (Miyuki Kawano 1999). Putative ORFs for both these genes were detected in all *E. hirae* ecotypes. However, an additional potassium transporting ATPase system comprising putative ATPase subunits A-C was detected only in *E. hirae-1* isolates. This 4260 bp region was also detected in the sequenced pig fecal *E. hirae-2* genomes but contained putative frameshifts in the subunit B gene suggesting that the operon is not functional in *E. hirae-2* isolates. This region was not detected in the *E.hirae-3* genomes. To determine the broader distribution of potassium transporting ATPase system, gene specific PCR assays targeting each subunit (A-C) were used. Positive PCR products were obtained for subunits A-C for 11/17 *E. hirae-1*, 13/17 *E. hirae-2* and 2/5 *E. hirae-3* isolates. All the positive PCR products were sequenced to confirm the identity of the PCR product. The expected frameshift in subunit B was detected in five *E. hirae-1*, all *E. hirae-2* and *E. hirae-3* isolates.

5.4.5.3 Copper homeostasis

Copper is an essential cofactor in respiratory, metabolic and stress response systems. Copper homeostasis is essential for balancing the effects of endogenous free radicals such as O_2^-. In *E. hirae*, copper uptake is regulated by the cop operon, which

contains a transcription regulator (CopY), a copper chaperone (CopZ), and two copper ATPases (CopA and CopB) (Solioz and Stoyanov 2003; Wunderli-Ye and Solioz 1999). A complete cop operon was detected in both of the *E. hirae-1* and *E. hirae-3* genomes. However, a frameshift resulting in deletion in the putative copB sequence was observed in both the *E. hirae-2* isolates. PCR targeting the copA gene yielded positive results in all the *E. hirae-1* and *E. hirae-2* (n=17) isolates and 4/5 *E. hirae-3* isolates. Similarly, positive PCR results for the putative copB gene were observed in all the *E. hirae-1* and *E. hirae-2* (n=17) and 4/5 *E. hirae-3* isolates. Sequencing of the positive PCR products for the putative copB sequence confirmed the presence of the expected frameshift in 11/17 *E. hirae-2* isolates and none of the *E. hirae-1* and *E. hirae-3* isolates.

A previously documented alternate mechanism for copper homeostasis was detected in *E. hirae-3* genomes. The cue operon includes a putative histidine kinase, a transcriptional regulatory protein (PhoB), a multi copper oxidase (blue copper oxidase CueO), a cation transport ATPase, a prolipoprotein diacylglyceryl transferase and five distinct ORFs for hypothetical proteins. In addition to the cue operon, another operon with a potential role in cellular copper homeostasis containing ORFs for a putative cAMP receptor protein (Crp) transcription regulator, a hypothetical protein, prolipoprotein diacylglyceryl transferase, a copper chaperone, and cation transporting ATPase was detected only in *E. hirae-3* isolates.

5.4.6 Putative antibiotic resistance genes

Resfinder v1.4 (http://cge.cbs.dtu.dk/services/ResFinder/), a web based tool for identification of acquired anti-microbial resistance genes was used to examine the *E.*

hirae ecotype genomes. Sequences for tetracycline resistance, *Tet (M)* and *Tet (L)*, were observed in *E. hirae-1* and *E. hirae-3* isolates. Lincosamide resistance gene sequence *lnu (B)* was observed only in *E. hirae-1*. No acquired anti-microbial resistance genes were detected in *E. hirae*-2 isolates. Putative resistance genes for aminoglycosides, beta-lactams, fluoroquinolone, fosfomycin, fusidic acid, phenicol, rifampicin, sulphonamide, trimethoprim and glycopeptide were not detected in any of the pig fecal isolates.

5.4.7 CRISPR sequences

Clustered regularly interspaced short palindromic repeats (CRISPRs) are DNA arrays that contain several unique sequences separated by short repeat sequences. Two types of CRISPR elements have been identified in bacteria: those associated with functional genes (CRISPR-cas) and those without associated genes (orphan CRISPRs). In our previous study both CRISPR-cas and orphan CRISPR were detected by PCR in *E. hirae* isolates from pig feces (Katyal et al. 2013). In this study, CRISPR associated sequences: cas1, cas2, csn1 and csn2 were detected in the *E. hirae-1* and *E. hirae-2* isolates, while only cas1 and cas2 genes were detected in *E. hirae-3*. In the current study, CRISPRfinder was used to identify CRISPR arrays in the pig fecal isolate genomes. At least one CRISPR array was identified in each *E. hirae-1* and *E. hirae-3* isolate but no CRISPR arrays were detected in *E. hirae-2* isolates. A BLASTn search for the orphan CRISPR sequences detected in these isolates by targeted PCR in our previous study did not reveal any significant matches. This is perhaps not surprising owing to the draft nature of the genomes since the structure of CRISPR arrays with multiple repeated sequences would present a challenge for the assembly algorithm used (Rho et al. 2012).

5.4.8 Diversity in phage sequences

Prophage sequences were identified in the pig fecal isolate genomes using PHAST. The proportion of each genome composed of predicted phage sequences ranged from 2.5% to 7.3% (Table 5.10), similar to what has been observed in genome sequences of *E. hirae* ATCC 9790 and seven *E. faecium* isolates (van Schaik et al. 2010). *E. hirae-3* genomes contained proportionately less DNA of phage origin than *E. hirae-1* and *E. hirae-2*, and correspondingly fewer unique prophage sequences (Table 5.11). Both *E. hirae-1* and *E. hirae-2* genomes contained the putative sequences for *Enterococcus* phage *phiEf11*, which was not detected in *E. hirae-3* genomes. Putative sequences for *Enterococcus* phage *EFAP-1* were found only in the *E. hirae-3* genomes.

Table 5.10 Number of putative prophage ORFs in the pig fecal *E. hirae* isolates and *E. hirae* type strain ATCC9790

***cpn*60 defined ecotype**	**Isolate**	**Number putative prophage ORFs (% of genome)**
E. hirae-1	57-9-G6	187 (6.8)
	78-9-C1	160 (5.7)
E. hirae-2	81-15-F4	178 (6.7)
	88-15-E9	188 (7.3)
E. hirae-3	57-3-H11	67 (2.5)
	67-3-C5	102 (3.8)
E. hirae-1	ATCC9790^{T}	194 (7.0)

Table 5.11 Putative prophage regions for each strain and *E. hirae* type strain

Isolate/ Region	Length (kb)	Completeness (score)[1]	Number of ORFs	Phage (NCBI accession ID)	GC (%)
57-9-G6					
1	36	Intact (150)	49	*Enterococcus* phage EFAP-1 (NC_012419)	32.77
2	12.8	Incomplete (50)	19	*Lactobacillus* phage J-1 (NC_022756)	33.22
3	43.6	Intact (150)	67	*Enterococcus* phage phiEf11 (NC_013696)	34.35
4	19.3	Incomplete (60)	30	*Listeria* phage A006 (NC_009815)	34.05
5	16.8	Incomplete (50)	22	*Listeria* phage B025 (NC_009812)	33.95
57-3-H11					
1	6.3	Incomplete (40)	9	*Clostridium* phage c-st (NC_007581)	36.48
2	41	Intact (150)	58	*Enterococcus* phage EFAP-1 (NC_012419)	38.67
67-3-C5					
1	49.2	Intact (150)	59	*Enterococcus* phage EFAP-1 (NC_012419)	37.25
2	23.2	Incomplete (30)	29	*Enterococcus* phage phiFL2A (NC_013643)	34.69
3	8.2	Incomplete (50)	14	*Lactococcus* phage bIL285 (NC_002666)	36.03
78-9-C1					
1	56.5	Intact (120)	51	*Listeria* phage A118 (NC_003216)	34.68
2	43.1	Intact (150)	45	*Enterococcus* phage EFAP-1 (NC_012419)	32.33
3	42.8	Intact (150)	64	*Enterococcus* phage phiEf11 (NC_013696)	34.28
81-15-F4					
1	44.8	Intact (120)	52	*Enterococcus* phage phiEf11 (NC_013696)	35.88
2	32.1	Questionable (90)	48	*Bacillus* phage phBC6A52 (NC_004821)	35.61
3	29	Intact (110)	40	*Listeria* phage A118 (NC_003216)	33.77
4	26.5	Questionable (70)	38	*Lactococcus* phage bIL285 (NC_002666)	34.69
88-15-E9					
1	55.3	Intact (150)	64	*Enterococcus* phage phiEf11 (NC_013696)	35.95
2	30.1	Intact (100)	46	*Bacillus* virus 1 (NC_009737)	33.22
3	43.2	Questionable (80)	43	*Bacillus* phage phBC6A52 (NC_004821)	34.7

Isolate/ Region	Length (kb)	Completeness (score)[1]	Number of ORFs	Phage (NCBI accession ID)	GC (%)
4	27.1	Intact (110)	35	Staphylococcus phage YMC/09/04/R1988 (NC_022758)	34.05
ATCC 9790^T					
1	28.3	Questionable (90)	29	*Listeria* phage B025 (NC_009812)	33.53
2	35.7	Incomplete (40)	43	*Enterococcus* phage phiFL3A (NC_013648)	33.73
3	50	Intact (120)	55	*Bacillus* phage BCJA1c (NC_006557)	34.69
4	38.6	Intact (140)	53	*Listeria* phage 2389 (NC_003291)	34.44
5	10.2	Incomplete (10)	14	*Paramecium bursaria* Chlorella virus 1(NC_000852)	30.18

[1] Completeness score: This score is calculated based on the prediction whether the prophage-like region contains an intact or incomplete prophage determined by the length and number of the sequences.

5.4.9 *In vitro* growth competition assays

Our comparative genomics results showed that *E. hirae-1* and *E. hirae-2* isolates have putative genes involved in transport of carbohydrates, amino acid metabolism (glutamine metabolism, proline metabolism and selenocystiene metabolism), potassium-transporting ATPases and copper homeostasis that were not detected in *E. hirae-3* isolates. When grown individually in a rich medium, TSB, no differences were detected in growth curves of ecotypes (results were combined for both isolates within each ecotype since no differences were detected between any pair) (Fig. 5.8 A). In fecal extract medium, although growth of all isolates was less than observed in TSB, both *E. hirae*-1 and *E. hirae*-2 isolates achieved a significantly higher OD_{600} values than isolates from *E. hirae*-3 (Fig. 5.8 B).

Based on these results, we hypothesized that isolates from *E. hirae-1* and *E. hirae-2* ecotypes would be more competitive that *E. hirae-3* isolates in co-cultures. We tested this hypothesis by determining the relative competitive fitness indices of *E. hirae* isolates from all ecotypes in *in vitro* growth competitions in pig fecal extract medium. Fecal extract medium was used as the growth medium as it mimics the natural source of isolation of the *E. hirae* isolates used in the study. Co-culture experiments were conducted for the 12 combinations of isolates illustrated in Table 5.12.

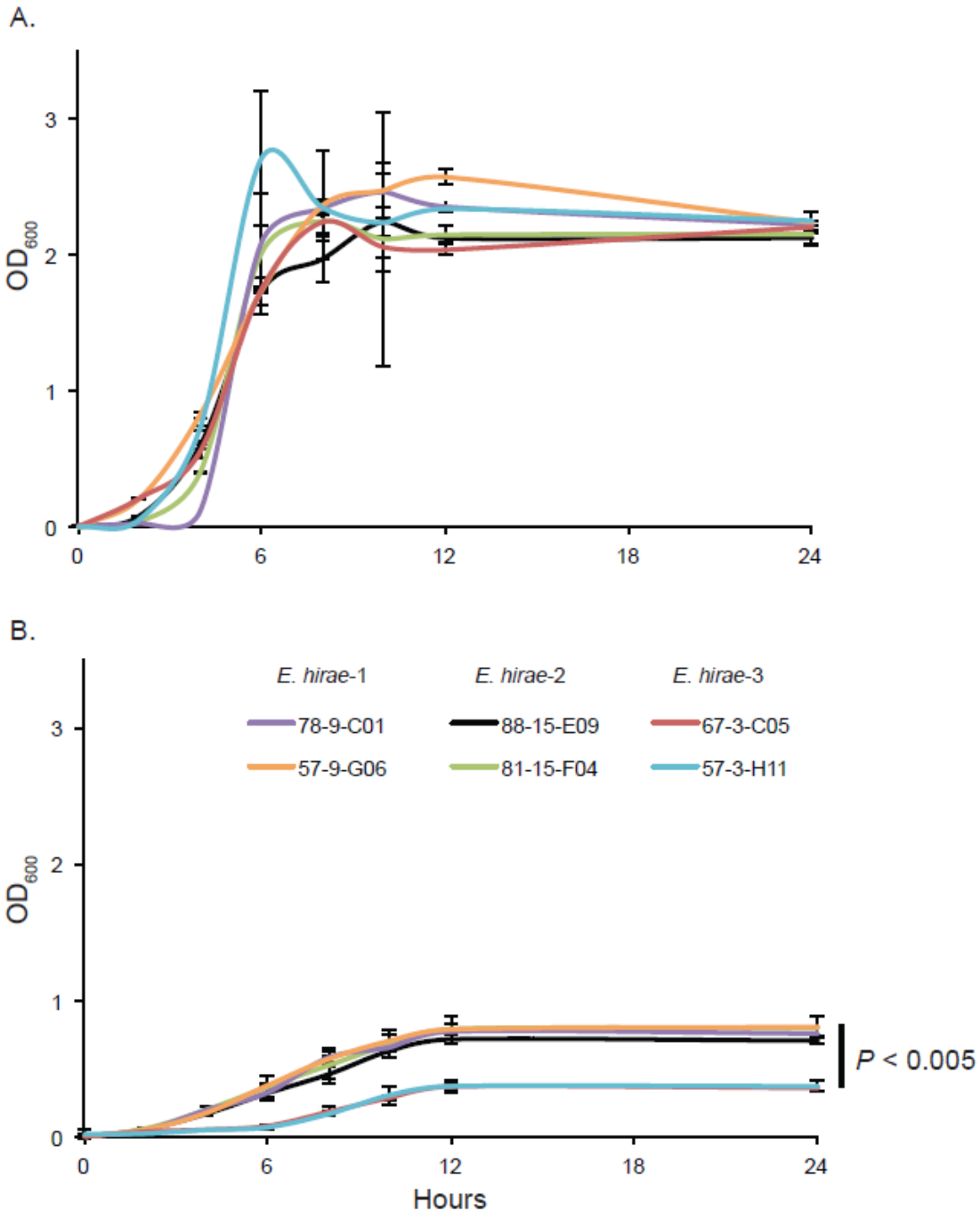


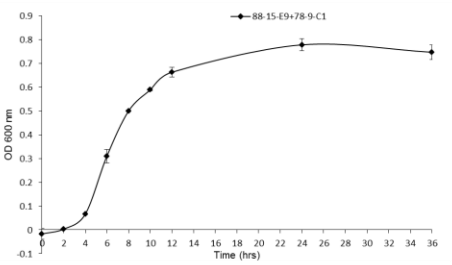
Figure 5.8 Growth curves of *E. hirae* isolates in (A) TSB and (B) fecal extract medium.
Values plotted are the mean of OD readings ± SD from biological triplicates from all the isolates within an ecotype. Kruskal-Wallis test was used to determine statistical significance. Error bars indicate standard deviation.

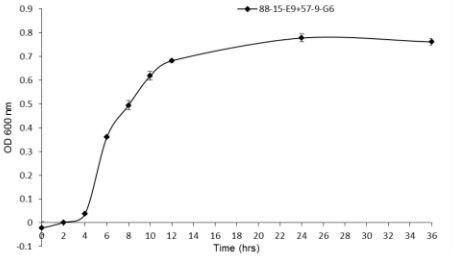
Table 5.12 Isolate combinations used in growth competition assays

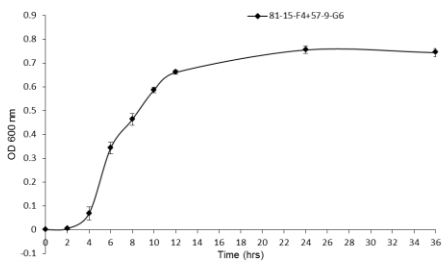
***cpn60* defined ecotype**	**Isolate**	**57-9-G6**	**78-9-C1**	**81-15-F4**	**88-15-E9**	**57-3-H11**	**67-3-C5**
E. hirae-1	57-9-G6			✓	✓	✓	✓
	78-9-C1			✓	✓	✓	✓
E. hirae-2	81-15-F4					✓	✓
	88-15-E9					✓	✓
E. hirae-3	57-3-H11						
	67-3-C5						

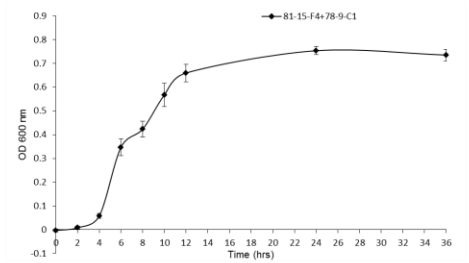
Growth curves for all mixed cultures were done in triplicate and the results are shown in Figure 5.9. No statistically significant differences between growth curves for the mixed cultures were observed, and the maximum OD_{600} values obtained were similar to those observed for individual culture (Figure 5.8 B). The abundance of each competing strain in the mixed cultures was determined at 24 hours by ecotype specific quantitative PCR assays (Figure 5.10). No differences in final concentration were detected between *E. hirae-1* and *E. hirae-2* isolates. Conversely, the final concentration of *E. hirae-3* isolates when grown in combination with *E. hirae-1* or *E. hirae-2* was significantly lower. A competitive fitness index (CFI) of 0.98 was observed for *E. hirae*-1 vs. *E. hirae*-2 indicating that the isolates are equally competitive in the pig fecal extract medium. CFI values for *E. hirae-2* and *E. hirae-1* isolates grown in competition with *E. hirae-3* in pig fecal extract were 1.51 and 1.75, respectively, indicating that *E. hirae*-1 and -2 both have competitive advantage over *E. hirae*-3.

A

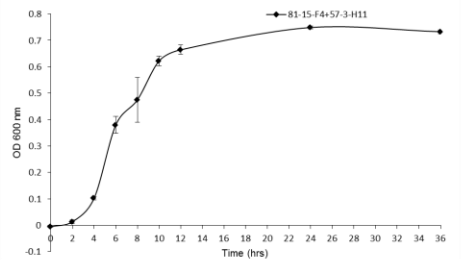

88-15-E9+78-9-C1
OD 600 nm
Time (hrs)

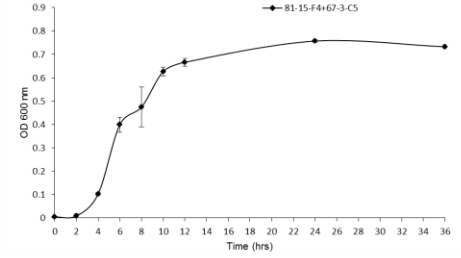

88-15-E9+57-9-G6
OD 600 nm
Time (hrs)

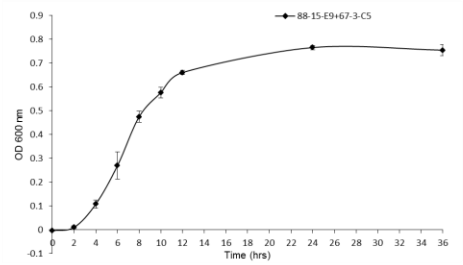

81-15-F4+57-9-G6
OD 600 nm
Time (hrs)

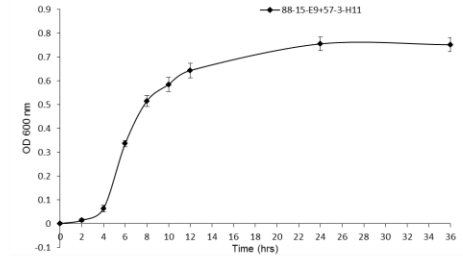

81-15-F4+78-9-C1
OD 600 nm
Time (hrs)

B

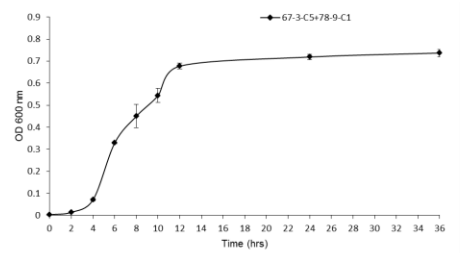

81-15-F4+57-3-H11
OD 600 nm
Time (hrs)

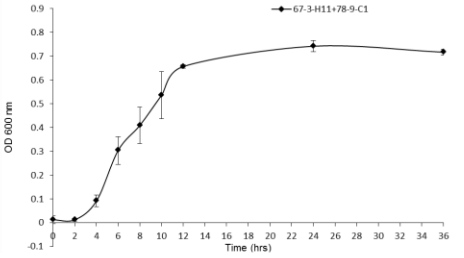

81-15-F4+67-3-C5
OD 600 nm
Time (hrs)

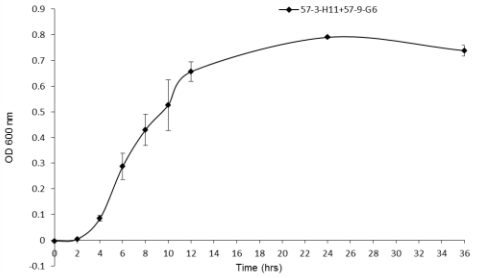

88-15-E9+67-3-C5
OD 600 nm
Time (hrs)


88-15-E9+57-3-H11
OD 600 nm
Time (hrs)

C


67-3-C5+78-9-C1
OD 600 nm
Time (hrs)


67-3-H11+78-9-C1
OD 600 nm
Time (hrs)


57-3-H11+57-9-G6
OD 600 nm
Time (hrs)

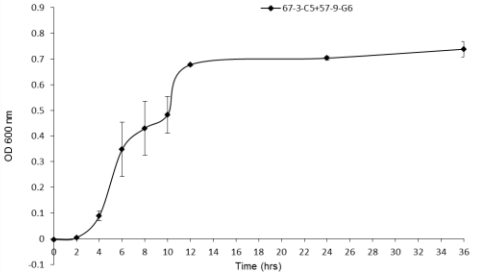

67-3-C5+57-9-G6
OD 600 nm
Time (hrs)

Figure 5.9 Growth curves for 12 competition experiments
a) *E. hirae-2* (81-15-F4 and 88-15-E9) vs. *E. hirae-1* (57-9-G6 and 78-9-C1) b) *E. hirae-2* (81-15-F4 and 88-15-E9) vs. *E. hirae-3* (57-3-H11 vs 67-3-C5) c) *E. hirae-3* vs. *E. hirae-1* (57-9-G6 and 78-9-C1). Values plotted are the mean of OD readings ± SD from biological triplicates.

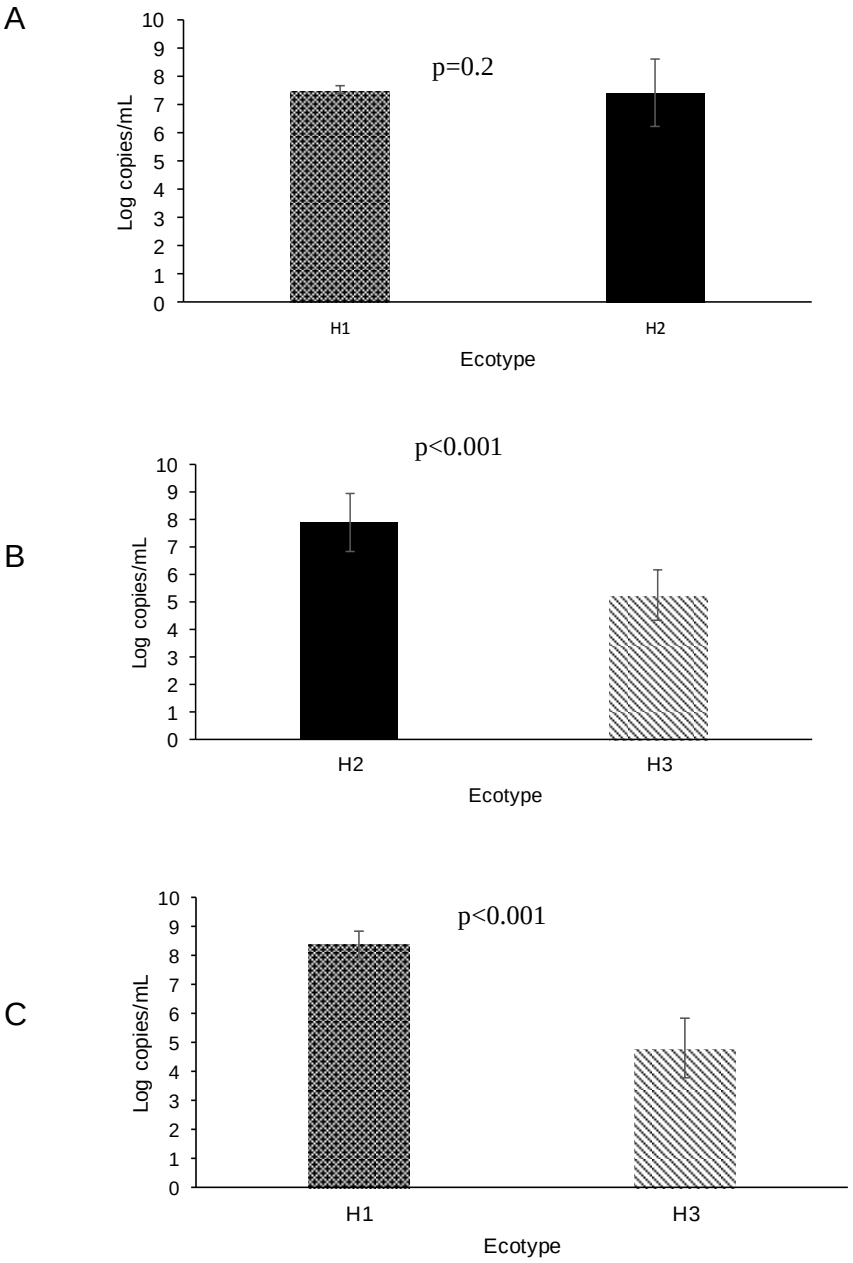


Figure 5.10 Abundance of each ecotype in the fecal extract medium at 24 hours measured by quantitative PCR

A: *E. hirae-2* (H2) vs. *E. hirae-1* (H1); *B: E. hirae-2* (H2) vs. *E. hirae-3* (H3), *C: E. hirae-3* (H3) vs. *E. hirae-1* (H1). Values shown are the mean of twelve OD readings ±SD from biological triplicates of both the isolates of each ecotype. Kruskal-Wallis test was used to determine statistical significance.

5.5 Discussion

Doolittle and Zhaxybayeva (2009) suggested that bacterial strains with "intermediate" phenotypes and genotypes may exist between species in natural environments. These intermediate populations may not be sampled either because of their low abundance or due to observations made on the basis of a specific marker gene that lacks the ability to resolve closely related strains within the community. Such "intermediate" strains have been recognized in the "microdiverse clusters" of sequences in large-scale metagenomic studies, which correspond to ecotypes: phenotypically and genotypically distinct strains that occupy distinct ecological niches (Cohan and Perry 2007).

In our previous study, we showed that phenotypically and genotypically distinct *E. hirae* ecotypes emerge through periodic selection in the pig fecal microbiome due to diet change, age and the physical location of the animal (Vermette et al. 2010). Resolution of ecotypes was based on *cpn*60 UT sequences and this resolution was not obtained using 16S rRNA sequence based methods. In this study, we compared the genome sequences of six *E. hirae* isolates (two from each *E. hirae* ecotype) to identify gene content differences that distinguish ecotypes, further defining the level of resolution revealed by *cpn*60 UT profiling of natural microbial communities.

Since the focus of the study was to identify gene content differences that could contribute to ecological functions in a microbial community we narrowed our search to genes with functional annotations. In our previous study (Vermette et al. 2010) and in the current study, it was seen that a difference in utilization of lactose as a sole carbon source exists between *E. hirae* ecotypes. In this study, we showed that *E. hirae-1* and *E. hirae-2*

isolates have genes that may be involved in acquisition of nutrients, maintenance of ion transport and metabolism of amino acids that are lacking in *E. hirae-3*. For example, a functional lactose specific PTS operon was found only in the sequenced *E. hirae-2* isolates. This operon was disrupted by a frameshift in *E. hirae-1* isolates and was not detected in *E. hirae-3* isolates. Differences in prevalence of the lactose specific EII CB fusion domain in *E. hirae-1* and *E. hirae-2* isolates generally correlated with the lactose utilization phenotype. However, 6/14 *E. hirae-1* isolates with a frameshift in the lactose specific EII CB domain were reported to utilize lactose as the sole carbon source. It is possible that these isolates have an alternate PTS EII domain that can transport lactose in addition to other lac family substrates. Francl et al. (2012) reported the presence of two EII domains, PTS 6CB and PTS 9BC, in *Lactobacillus gasseri* ATCC 33323, suggesting that multiple EII complexes may be involved in transporting lactose. However, an EII BC fusion domain was not detected in the sequenced *E. hirae* isolates. In addition, a lac Y gene was not observed in any of the sequenced *E. hirae* isolates. This is not surprising as Kandler (1983) suggested that streptococci (the genus *Enterococcus* was classified as group D streptococci until 1984) rely on a PTS to transport lactose or have cryptic lac Y and β-galactosidase genes. Further, studies in several species of *Streptococcus*, *Lactobacillus* and Firmicutes in general have shown that a lactose uptake is PTS dependent (Cases et al. 2007; Francl et al. 2012; Francl et al. 2010; McKay et al. 1970; Thompson 1979). Crow et al. (1983) have shown that the PTS system in *Streptococcus lactis* is plasmid associated and thus, diversity of the lactose specific PTS operon in the sequenced *E. hirae* isolates could be attributed to association of this domain with a mobile genetic element.

Our study provides the first description of PTS in *E. hirae*, specifically the lac family PTS. Putative genes for all the PTS components (EI, HPr, EII A, EII B, EII C and EII D) were found in all the sequenced isolates, which is consistent with what is known for Firmicutes (Cases et al. 2007). The total percentage of ORFs encoding PTSs in the sequenced *E. hirae* genomes (2.2%-2.7%) is consistent with previous observations that up to 3.2% of bacterial genomes encode PTS (Barabote and Saier 2005). All seven families of putative EII permeases were found in the sequenced genomes. The abundance of seven EII permease families in the sequenced *E. hirae* genomes is generally consistent with other Firmicutes where the Glc family EII components are also the most abundant. Interestingly, the number of components for each EII family in *E. hirae* isolates is greater than that reported for *E. faecalis* V583 and other members of order *Lactobacillales* (Barabote and Saier 2005). The order of abundance of EII components for Lac, Man and Fru family PTS in the sequenced *E. hirae* genomes is different from what has been observed for the order *Lactobacillales* where the number of Man family PTS was more than Lac and Fru family PTS components. It is important to note that the genomic sequences of strains (n=106) used in the comparative genomic study by Barabote and Saier (2005) were obtained from publically available databases that are dominated by pathogenic strains and thus a higher abundance of Man family PTSs that are involved in establishment of relationships between bacteria and their eukaryotic hosts is not surprising (Comas et al. 2008; Zuniga et al. 2005). In our case though, a higher abundance of lac family PTS EII component could be due to niche adaptation of *E. hirae* isolates in the GI tract.

In addition to the diversity in PTS, we have shown that putative genes involved in glutamine and proline metabolism were present only in *E. hirae-1* and *E. hirae-2* isolates and were not detected in *E. hirae-3* isolates. The absence of amino acid biosynthesis pathways has been linked to niche specialization in the case of *Lactobacillus bulgaricus* (the hallmark bacterium found in yogurt), reflecting the bacterium's adaptation to the protein-rich milk-based diet (Pfeiler and Klaenhammer 2007).

The operon for selenocystiene biosynthesis was present only in isolates of *E. hirae-1* ecotype. Selenocystiene metabolism genes have been reported in Firmicutes (Clostridia, Mollicutes and Bacillales*)*, but *Enterococcus faecalis* is the only member of the Lactobacillales family to date that has been reported to contain putative genes for selenocystiene biosynthesis (Zhang et al. 2006). Selenocystiene is present in the active sites of seleno-enzymes that are involved in redox function in archaea, bacteria and eukaryotes (Stadtman 1996). Hatfield and Gladyshev (2002) suggested that replacement of cysteine with selenocystiene in seleno-enzymes provides a selective advantage to an organism by enhancing the efficiency of seleno-enzymes in comparison to their cysteine containing homologs. Zhang et al. (2006) suggested that organisms possessing genes for selenocystiene synthesis are found in anaerobic environments, such as the distal gastrointestinal tract. Factors determining loss and acquisition of genes involved in selenocystiene biosynthesis are not completely understood but the involvement of lateral gene transfer and the presence of selenium in the environment have been suggested (Zhang and Gladyshev 2008; Zhang et al. 2006). Thus, presence of putative genes for selenocystiene biosynthesis in *E. hirae-1* could be indicative of a changing GI environment due to dietary changes and niche adaption by this ecotype.

Competition is one the most important interactions in microbial communities as all microbes need nutrition and space for growth. It plays a significant role in shaping community composition, as the microbes that are the "best fit" for an environment will out-compete others. Fitness can be defined as the ability of a bacterium to grow and reproduce in an environment and depends on factors such as biosynthetic capacity, ability to produce anti-microbial compounds, and virulence (Orr 2009). One of the most widely accepted ways to measure fitness is to use mathematical models to compare growth of individual strains in mixed cultures and determine the CI. Simply, CI is defined as the ratio of quantity of one strain to another within the output sample, divided by the corresponding ratio in the inoculum (Travisano 1997). To date, most studies to determine the fitness (expressed as CI) of a strain have been done using pathogenic/non-pathogenic and/or drug resistant/susceptible bacteria (Baumler et al. 1997; Bhatter et al. 2012; Freter and O'Brien 1981; Macho et al. 2007; Monk et al. 2008) and results of these studies have shown that competing strains can affect each other's growth rate and ability to utilize available nutrients, with the more fit strains outgrowing the less fit strains. The fitness of any strain in a given environment will also contribute to its success in a community since even without any direct interaction, a slower growing strain will be at a competitive disadvantage relative to neighbours with higher fitness.

We compared the competitive fitness indices of *E. hirae* ecotypes by using pig fecal extract medium to simulate the *in vivo* conditions from which these isolates were originally obtained. In individual culture experiments, *E. hirae-1* and *E. hirae-2* isolates had greater fitness than the *E. hirae-3* isolates in the pig fecal medium as predicted by the genomic comparisons. This is consistent with *E. hirae-1* and *E. hirae-2* isolates having

higher biosynthetic capacities than *E. hirae-3* isolates and can survive in relatively nutritionally deficient medium such as the fecal extract that may also contain secondary metabolites, feces associated phages and other cellular molecules such as host associated DNA/RNA. In co-culture experiments, *E. hirae-1* and *E. hirae-2* achieved higher culture densities than *E. hirae-3*. These results were perhaps predictable based on the growth of individual isolates in fecal extract. In addition to competition for nutrients and space, the possibility of anti-microbial compounds such as bacteriocins cannot be ruled out but our experimental design did not allow us to draw conclusions about direct competitive interactions. However, our results do suggest that there was no gain in competitive fitness conferred on the less fit *E. hirae-3* by culture with either of the other ecotypes. Results from our competition assays may explain the rareness of *E. hirae-3* ecotype in the pig fecal *Enterococcus* community, but the shift in community composition between 9 and 15 weeks from *E. hirae-1* to *E. hirae-2* dominance could not be explained in our current study. This may be due to the fact that the pig fecal extract medium was prepared from pigs that were seven weeks old and was not suitable to tease apart the differences in competitiveness of *E. hirae-1* and *E. hirae-2* isolates.

In conclusion, we show that genome level diversity exists between *E. hirae* ecotypes and have identified several specific gene content differences that are likely related to niche specialization. The gene content differences observed were related to differential biosynthetic capacities and competitive fitness differences observed between *E. hirae* ecotypes, potentially explaining the succession observed in the pig fecal *Enterococcus* community. Our results provide further evidence that the *cpn*60 UT provides a high-resolution tool for studying microbial community dynamics that allows

for subspecies resolution and identification of phenotypically and genotypically distinct ecotypes in complex bacterial communities.

APPENDIX

Ecotype specific primer design for real time PCR assays

Genome sequences from each draft genome sequence were compared to another by reciprocal reference mapping. Reciprocal reference mapping where reads from one genome were matched to reads from another and vice-versa was performed between all *E. hirae* isolates using gsMapper v2.7. Unmapped reads from each reciprocal pairwise comparison were assembled into contigs to identify ecotype-specific genome regions. Primer sequences for each ecotype were designed to target DNA sequences unique to that ecotype. Primers were designed using Beacon Designer (Premier Biosoft, Palo Alto, CA). Specificity of each primer set was checked *in silico* by conducting a BLAST search for each primer against all the genome sequences. Optimal annealing temperatures for all primer sets was determined by performing a temperature gradient PCR with annealing temperature ranging from 49.8°C to 70°C. Specificity of each primer set was evaluated by testing each primer set by using genomic DNA from all the pig fecal *E. hirae* isolates separately in both the conventional and real-time PCR reactions as described below. Each assay was optimized for ecotype specificity by using a temperature gradient PCR and evaluating the amplification curves, dissociation peaks and efficiency of the reactions. The list of primer sequences, expected product size and annealing temperature for each primer set is given in Table A. 1.

Table A.1 Primer sequences and optimal annealing temperatures for ecotype-specific assays

***cpn*60 defined ecotype**	**Primer sequence (5'-3')**	**Expected product size (bp)**	**Annealing temperature for real time PCR (°C)**
E. hirae-1	Forward - ATACTTGAATCATCTTCGTAT	227	58.0
	Reverse - GTAGGAGGTCTTATTACTTATAG		
E. hirae-2	Forward - CAGTTGCTAATTCTTCACCTATTG	269	55.0
	Reverse - AGTTATCCAGTCGCTCCTTA		
E. hirae-3	Forward - TCATCAGCCACATATCTTCT	114	61.9
	Reverse - TGACTTACCACAGCCATT		

Conventional PCR for ecotype specific assays

For conventional PCR, each 50 µl reaction mixture contained 1 × PCR reaction buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 2.5 mM $MgCl_2$, 50 pmol of each primer (Table A. 1), 0.2 mM dNTPs (Invitrogen, Burlington, ON), and 2.5 units Taq polymerase (Quanta Biosciences, Gaithersburg, MD). PCR amplifications were done in an Eppendorf Mastercycler EP gradient thermocycler with an initial denaturation step (95ºC for 3 minutes), followed by 40 cycles of denaturation (95ºC, 30 seconds), annealing (temperature in Table 5.1, 30 seconds) and extension (72°C, 1 minute), and a single final extension step (72ºC, 10 minutes). Results of PCR reactions were evaluated by resolution of the products on 1.5% agarose gels stained with ethidium bromide, and visualized and photographed under UV light. Amplified products were purified using EZ-10 spin column PCR purification kit (Bio Basic Inc., Markham, ON) and sequenced using their respective PCR primers (Table A.1). The specificity of ecotype specific assays using conventional PCR is shown in Figure A.1.

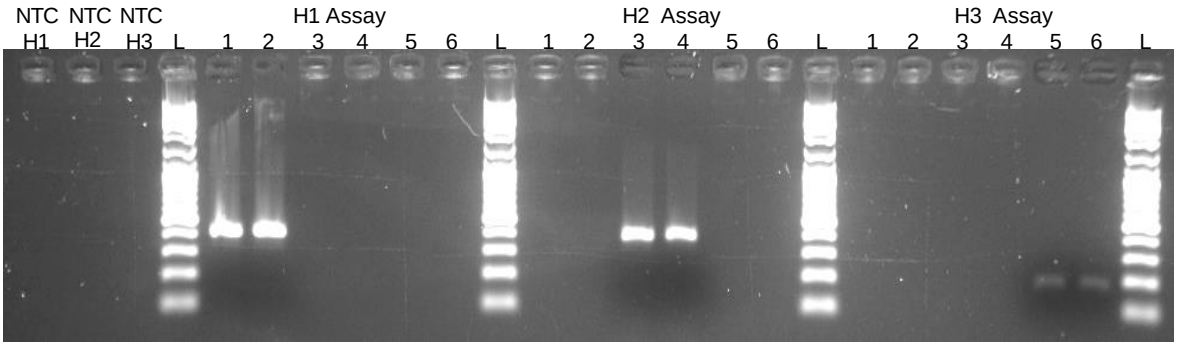
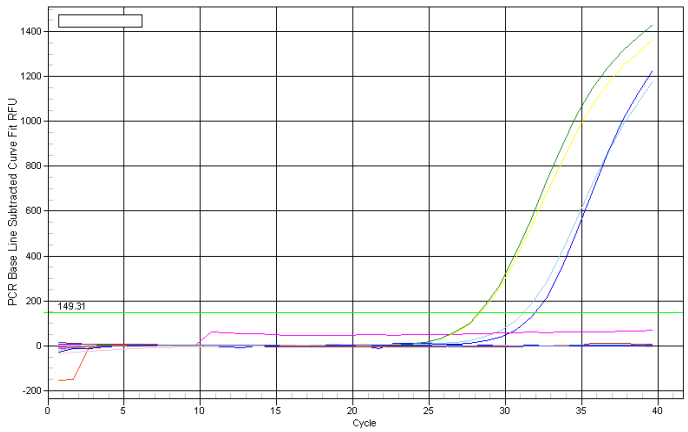


Figure A.1 *E. hirae* ecotype specific PCR products from *E. hirae-1* (H1), *E. hirae-2* (H2) and *E. hirae-3* (H3) isolates.
Lanes are (NTC) PCR no template control; (L) Ladder=10,000-100 bp (1) *78-9C1 (E. hirae-1)*; (2) 57-9-G6 (*E. hirae-1*), (3) 81-15-F4 (*E. hirae-2*), (4) 88-15-E9 (*E. hirae-2*), (5) 57-3-H11 (*E. hirae-3*), (6) 67-3-C5 (*E. hirae-3*).

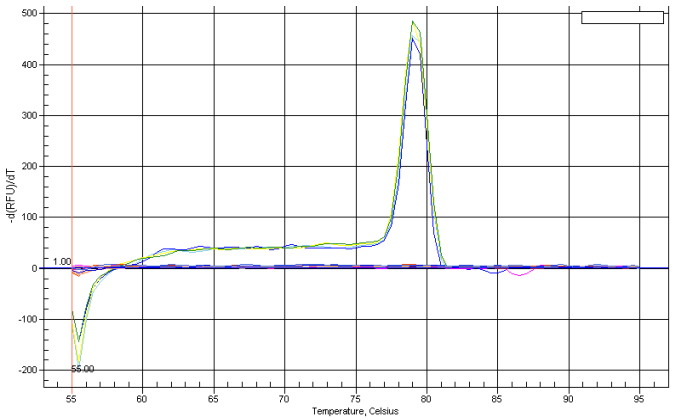
Real time PCR assays

All real-time quantitative PCR (qPCR) reactions were run with a no template control (NTC) and a standard curve composed of plasmids containing ecotype specific target sequences added at concentrations of 10^0 to 10^7 (10 fold series dilution) plasmid copies/reaction. Positive control plasmids containing PCR products for each ecotype specific target were constructed by cloning the respective PCR products into pGEM-T-Easy vector (Promega, Madison, WI). All qPCR reactions were done in duplicate. Each reaction consisted of 1 × iQ SYBR Green supermix (BioRad), 400 nM each of primer (Table A. 1) and 2 μL of template DNA, in a final volume of 25 μL. A thermocycler (MyiQ; BioRad, Mississauga, ON) was used for all reactions with the following program: 95°C for 3 min, followed by 37 cycles of 95°C for 15 s, specific annealing temperature as listed in Table A. 1 for 15 s, 72°C for 15 s, and a final extension at 72°C for 5 min. A melt curve was subsequently run for 81 cycles at 0.5°C increments from 55°C to 95°C for 30 s at each time point. Fluorescent signals were measured every cycle at the end of the annealing step and continuously during the melt curve data collection. The data obtained was analyzed using iQ5 Optical System Software (BioRad, Mississauga, ON). A sample was classified as positive if both technical replicates were positive and had a Ct standard deviation of less than 1. The results for real time PCR assays are shown in Figures A.2-A.4.

A



B



C

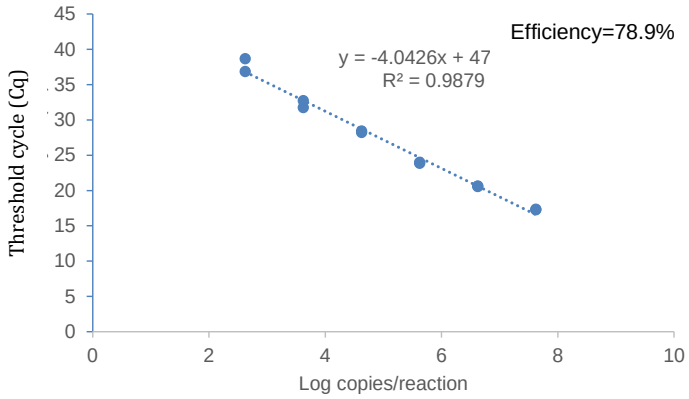
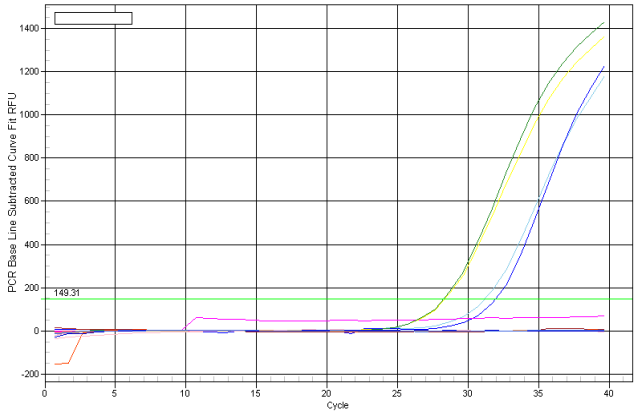
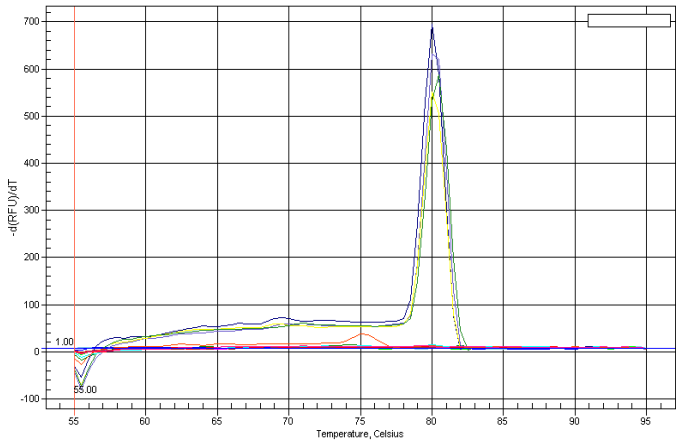


Figure A.2 Real time PCR assay for *E. hirae-1*
a) Amplification curve and (b) melt peak for analytical specificity of *E. hirae-1* specific SYBR green real time PCR assay. (c) Standard curve plot (copy number vs. threshold cycle (C_q) with regression and efficiency) for *E. hirae-1* specific DNA in plasmid (positive control) detected by SYBR green real-time PCR. Individual data points represent duplicate wells containing 25 µL reactions.

A



B



C

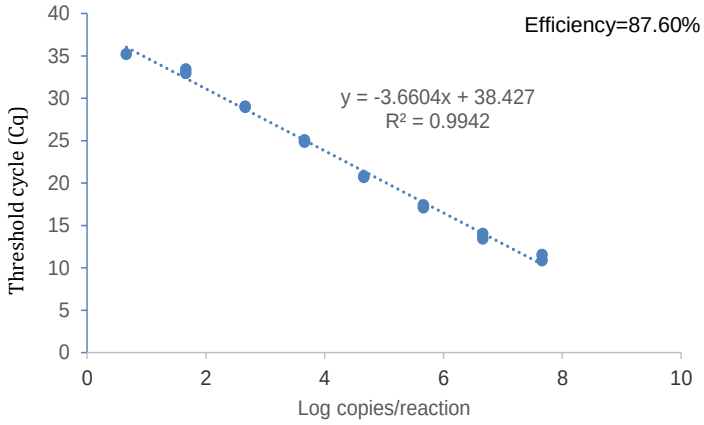
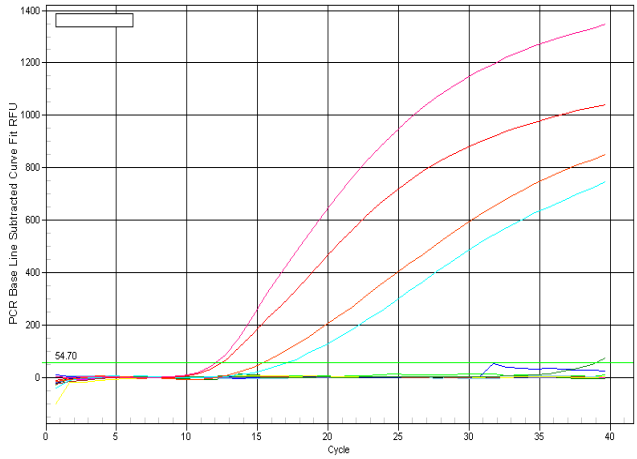


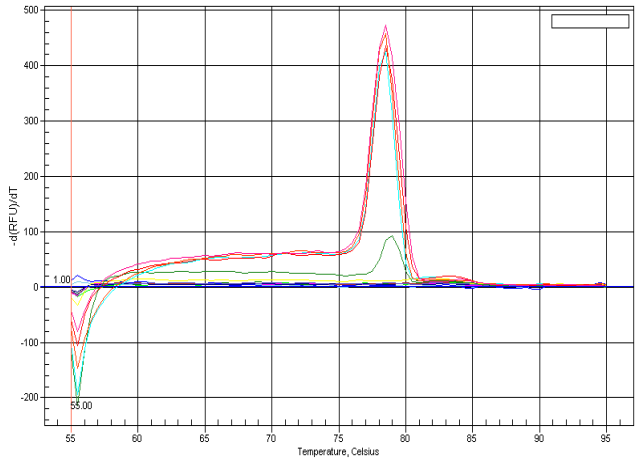
Figure A.3 Real time PCR assay for *E. hirae-2*

(a) Amplification curve and (b) melt peak for analytical specificity of *E. hirae-2* specific SYBR green real time PCR assay. (c) Standard curve plot (copy number vs. threshold cycle (C_t) with regression and efficiency) for *E. hirae-2* specific DNA in plasmid (positive control) detected by SYBR green real-time PCR. Individual data points represent duplicate wells containing 25 µL reactions.

A



B



C

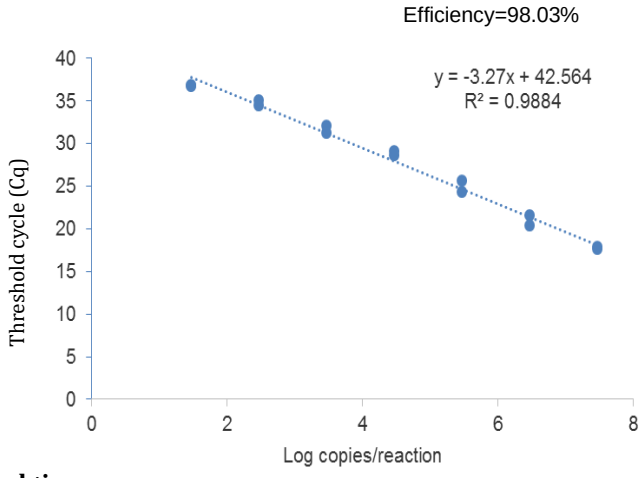


Figure A.4 Real ti

(a) Amplification curve and (b) melt peak for analytical specificity of *E. hirae-3* specific SYBR green real time PCR assay. (c) Standard curve plot (copy number vs. threshold cycle (C_t) with regression and efficiency) for *E. hirae-3* specific DNA in plasmid (positive control) detected by SYBR green real-time PCR. Individual data points represent duplicate wells containing 25 µL reactions.

CHAPTER 6 - General discussion and conclusions

6.1 Summary and limitations of these works

6.1.1 *cpn*60 is a preferred tool for identification of *Enterococcus* species and sub-species

The utility of *cpn*60 UT based methods for distinguishing *Enterococcus* species was first established in the year 2000 and the analysis included 19 *Enterococcus* species (Goh et al. 2000). Since the year 2000, the total number of species has increased to 49 (at the time of writing) (Parte 2014) and includes species from diverse habitats. Simultaneously, the number of *Enterococcus* related hospital acquired infections has increased tremendously. Out of all known *Enterococcus* species, *E. faecium* has emerged as one of the leading causative agents of multidrug resistant infections in the US (Agudelo Higuita and Huycke 2014; Lebreton et al. 2013). Although, several biochemical and molecular detection methods are available, accurate identification and detection of *Enterococcus* species is an unmet need in clinical microbial settings (Pendle et al. 2008).

The results presented in this thesis highlight the usefulness of the *cpn*60 UT sequence for accurate identification of *Enterococcus* species. In Chapter 2, we demonstrated the usefulness of using *cpn*60 UT sequences for resolving *Enterococcus* species, including *Enterococcus* species that have been discovered since 2000. Using *cpn*60 UT sequences, we could distinguish all of the 28 *Enterococcus* species for which *cpn*60 sequences are available. It is important to note here that this resolution was not possible using 16S rRNA gene sequences. This is not surprising as Ryu et al. (2013) observed cross-reactivity of an *Enterococcus* group specific qPCR assay with non-

targeted species. The authors proposed that 16S rRNA based method should be used together with biochemical tests for accurate identification of *Enterococcus* species. Although other genes such as *tufA, sodA, rpoB and pheS* (Jackson et al. 2004; Ke et al. 1999; Naser et al. 2005a) have been investigated for identification of *Enterococcus* species, these methods are limited by the lack of a comprehensive sequence database and universal PCR primers.

We also utilized *cpn*60 UT sequences to determine if sub-species level resolution could be achieved for *E. faecium* and *E. faecalis*. It was shown that *E. faecium cpn*60 UT sequences could be resolved into two distinct clades, which correspond with the results from multilocus sequence typing (MLST) and comparative genomics studies of *E. faecium*. These results provide evidence for applicability of the *cpn*60 UT as a stand-alone tool that can be used for identification of *E. faecium* isolates from diverse environments.

A limitation of these works is that the analysis could not be validated for all the known *Enterococcus* species due to unavailability of *cpn*60 UT sequence for 20 *Enterococcus* species.

6.1.2 GTG rep PCR shows that genomic diversity exists within *cpn*60 ecotypes

In a previous study in our lab, a shift in the fecal *Enterococcus* community composition in healthy pigs with change in diet over time was seen. It was also shown that *cpn*60 UT sequences could resolve ecotypes (genotypically and phenotypically distinct strains within a species) within *E. faecalis* and *E. hirae*. However, the genotypic distinction in the study by Vermette et al. (2010) was based on the 552 bp *cpn*60 UT

sequences and thus provided only limited information about the level of genotypic diversity between and within ecotypes of *E. hirae* and *E. faecalis*.

In Chapter 3, we used GTG rep PCR to detect genomic diversity among pig fecal *Enterococcus* ecotypes. Our results show that GTG rep PCR is a useful technique to detect genome level diversity between and within *Enterococcus* ecotypes. This diversity could be due to genome re-arrangements and/or genome content differences. Our results further emphasize the usefulness of the *cpn*60 UT sequences to predict genome relatedness as observed in previous studies (Verbeke et al. 2011) since clustering of isolates by GTG rep PCR fingerprints was generally consistent with the *cpn*60 UT phylogeny.

A limitation of this work is the lack of *Enterococcus* isolates from multiple animals, thereby limiting the analysis of genomic diversity in isolates from only a few animals. Also, although the fingerprinting technique is useful for detecting difference in genome structure and presumably content that may be useful for molecular epidemiological studies, no detailed information on gene content differences between ecotypes is generated.

6.1.3 CRISPR analysis provided a snapshot of inter-specific interactions in the *Enterococcus* community

Mobile genetic elements such as phages, anti-microbial resistance genes and pathogenicity islands may contribute to shaping the microbial community composition in animals. Moreover, the GI tract has been considered a “hot-spot” for exchange of genetic material between microbes. CRISPRs can be used as records of microbe–virus

interactions in animal microbial communities and may provide insights into previously unrecognized factors that play a role in determining succession in bacterial communities. With this information and the availability of primers for detecting *E. faecalis* CRISPR loci, we have described the diversity in CRISPR sequences in the pig fecal *Enterococcus* community in Chapter 4. Out of all the spacer sequences identified in CRISPR arrays, only 40% were shared by all the pig fecal *E. hirae* and *E. faecalis* isolates. In addition, both species had distinct repeat sequences. Only two spacers matched previously identified pig viral sequences while others had similarity to *Enterococcus* genome sequences. Our results provide further evidence that mechanisms for integration of spacers in *E. faecalis* and *E. hirae* may be distinct, and that CRISPRs record exposure to other bacteria within the microbial community as well as viral pathogens.

A limitation of this work is that primers that were designed for *E. faecalis* isolates were used for screening the pig fecal *E. hirae* isolates. This could be an important factor in the lower abundance of CRISPR loci that were detected in *E. hirae* isolates. A major reason for the lack of primers for detection of CRISPR loci in *E. hirae* is the unavailability of complete genome sequences of *E. hirae*. Future work to sequence genomes of *E. hirae* isolates could facilitate identification of CRISPR loci and aid in development of species-specific primers. Knowing the palindromic nature of CRISPR loci, sequencing methods that generate shorter reads and guided sequence assembly methods could aid in identification of CRISPR loci in *E. hirae* genomes. In addition, use repeat sequences and/or conserved cas gene loci to design primers and sequence the CRISPR arrays in *E. hirae* could be an alternative.

6.1.4 Comparative genomics of pig fecal *E. hirae* ecotypes reveals ecotype specific genes involved in niche specialization and competitive fitness

Results described in the previous chapters indicated that phenotypically and genotypically distinct ecotypes exist within *E. faecalis* and *E. hirae* communities obtained from feces of healthy pigs. Also, a shift in community composition was observed with change in diet and age of the animals. However, no information about the depth of genomic diversity was obtained. Doolittle and Zhaxybayeva (2009) proposed that strain specific genes play a critical role in niche specialization and bacterial community succession. In Chapter 5, we showed that gene level diversity exists within ecotypes and includes genes involved in niche specialization. The work provided evidence that the *cpn*60 UT is a useful tool to identify ecotypes in natural bacterial communities and highlights the kind of differences that can exist in ecotypes. More importantly, the work presented here shows that the micro-heterogeneity observed in single gene based phylogenetic studies is biologically relevant if the right tools such as the *cpn*60 UT are used. Results from the growth competitions between *E. hirae* ecotypes in pig fecal extract confirmed that the biosynthetic capacity of *E. hirae*-1 and *E.hirae*-2 isolates is greater than *E. hirae*-3 isolates.

A limitation of this work is that genome content differences between ecotypes could not be identified beyond genes for which functional annotations are available. It is a possibility that we are missing out on characterization of other relevant genes that play a significant role in determining the succession patterns in a bacterial community but have no known functional annotation. Focused studies for functional characterization of

these ecotype specific genes will greatly enhance our understanding of genomic diversity in dynamic and diverse bacterial communities. In addition, work presented in Chapter 5 regarding the differences in biosynthetic capacities between *E. hirae* ecotypes was obtained by conducting growth completion in fecal extract medium. Although use of pig fecal extract as a medium for growth completion is justified as it is a medium with close resemblance to the environment from which these isolates were initially obtained, lack of information regarding the composition of the fecal extract medium limits our ability to obtain detailed information about the biosynthetic capacity of each ecotype. Growth in minimal media could have been an alternative to compare the biosynthetic capacity of *E. hirae* ecotypes.

6.2 Discussion of future prospects

6.2.1 Obtain *cpn*60 UT sequences for the remaining 20 *Enterococcus* species and further validate the resolving power

The ubiquitous nature of *Enterococcus* species and emergence of vancomycin-resistant enterococci (VRE) emphasize the need for rapid and accurate detection of *Enterococcus* species. In Chapter 2, the ability of *cpn*60 UT to resolve *Enterococcus* species and *E. faecium* sub-species was demonstrated. Availability of *cpn*60 UT sequences for the remaining 20 *Enterococcus* species will allow validation of utility of *cpn*60 UT for development of diagnostic assays. Molecular methods based on PCR and sequencing of the *cpn*60 UT or multiplex PCR assays will facilitate rapid and accurate detection of *Enterococcus* species in complex microbial communities and clinical

samples. Moreover, commercially available kits for detection of *Enterococcus* rely on phenotypic characteristics and include use of selective media and may take between 24 to 48 hours. Molecular methods that have been used for detection of *Enterococcus* at the genus level use 16S rRNA gene and provide limited resolution at the species level. Species-specific PCR based kits for detection and identification of *Enterococcus* species are available but are limited to a few clinically relevant species including *E. faecalis, E. faecium* and *E. casseliflavus*. The utility of *cpn60* UT based PCR assays has been validated for accurate detection of several species including *Streptococcus*, *Staphylococcus* and *Campylobacter* (Goh et al. 1998; Goh et al. 1997; Hill et al. 2006a). Also, the *cpn*60 UT has been used for development of multiplex assays for detection and differentiation of Phytoplasmas (Dumonceaux et al. 2014) and for characterization of vaginal microbial communities (Dumonceaux et al. 2011). Development of *cpn*60 UT based methods for identification of *Enterococcus* species and sub-species will results in better diagnostic tools for both diagnostic laboratories and hospitals.

6.2.2 Exploration of complex microbial communities at a level beyond species

The current state-of-art for studying microbial community composition and succession is the use of 16S rRNA based pyrosequencing methods (Cole et al. 2007). There is ample evidence in literature that describes the limited ability of 16S rRNA based sequencing methods for study of complex microbial communities. Moreover, the microdiversity observed in the phylogenetic analysis based on 16S rRNA sequences is often ignored as PCR artifacts.

The results presented in this thesis provide evidence that it is important to consider microdiversity in studies of microbial community succession. With the use of *cpn*60 UT based methods, phenotypically and genotypically distinct *Enterococcus* ecotypes in complex microbial community associated with pig feces could be identified. Gene level diversity between the *E. hirae* ecotypes suggests that the micro-diversity observed with high resolution tools such as the cpn60 UT are biologically relevant and should not be ignored as PCR artifacts. In this study, we showed that the *cpn*60 UT is a useful tool for identification of the *E. hirae* ecotypes that are adapted to environments that differ in their nutrient composition. In addition, we have identified differences in distribution of putative genes that likely play a role in determining the ecological function of ecotypes. Our results could be of importance for studies aimed at defining bacterial species and investigating the process of speciation in complex microbial communities. Future studies to understand the usefulness of the *cpn*60 UT for understanding the dynamics of microbial communities obtained from diverse environments should be conducted.

REFERENCES

Aakra, A., Nyquist, O.L., Snipen, L., Reiersen, T.S., and Nes, I.F. 2007. Survey of genomic diversity among *Enterococcus faecalis* strains by microarray-based comparative genomic hybridization. Appl Environ Microbiol **73**(7): 2207-2217.

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol **71**(12): 8966-8969.

Adekambi, T., Drancourt, M., and Raoult, D. 2009. The rpoB gene as a tool for clinical microbiologists. Trends Microbiol **17**(1): 37-45.

Adekambi, T., Shinnick, T.M., Raoult, D., and Drancourt, M. 2008. Complete rpoB gene sequencing as a suitable supplement to DNA-DNA hybridization for bacterial species and genus delineation. Int J Syst Evol Microbiol **58**(Pt 8): 1807-1814.

Agudelo Higuita, N.I., and Huycke, M.M. 2014. Enterococcal Disease, Epidemiology, and Implications for Treatment. *In* Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. *Edited by* M. S. Gilmore, D. B. Clewell, Y. Ike, and N. Shankar, Boston.

Allen, H.K., et al. 2011. Antibiotics in feed induce prophages in swine fecal microbiomes. MBio **2**(6): e00260-00211.

Andrewes, F.W., and Horder, T.J. 1906. A study of streptococci pathogenic for man. Lancet **168(4335)**: 852–855.

Aoi, Y., Kinoshita, T., Hata, T., Ohta, H., Obokata, H., and Tsuneda, S. 2009. Hollow-fiber membrane chamber as a device for in situ environmental cultivation. Appl Environ Microbiol **75**(11): 3826-3833.

Aziz, R.K., et al. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**: 75.

Baele, M., Devriese, L.A., Butaye, P., and Haesebrouck, F. 2002. Composition of enterococcal and streptococcal flora from pigeon intestines. J Appl Microbiol **92**(2): 348-351.

Bakker, E.P., and Harold, F.M. 1980. Energy coupling to potassium transport in Streptococcus faecalis. Interplay of ATP and the protonmotive force. J Biol Chem **255**(2): 433-440.

Barabote, R.D., and Saier, M.H., Jr. 2005. Comparative genomic analyses of the bacterial phosphotransferase system. Microbiol Mol Biol Rev **69**(4): 608-634.

Barrangou, R., et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315**(5819): 1709-1712.

Baumler, A.J., Tsolis, R.M., Valentine, P.J., Ficht, T.A., and Heffron, F. 1997. Synergistic effect of mutations in invA and lpfC on the ability of *Salmonella typhimurium* to cause murine typhoid. Infect Immun **65**(6): 2254-2259.

Ben Omar, N., et al. 2004. Functional and safety aspects of *Enterococci* isolated from different Spanish foods. Syst Appl Microbiol **27**(1): 118-130.

Berga, M., Szekely, A.J., and Langenheder, S. 2012. Effects of disturbance intensity and frequency on bacterial community composition and function. PLoS One **7**(5): e36959.

Berrington, J.E., Stewart, C.J., Embleton, N.D., and Cummings, S.P. 2013. Gut microbiota in preterm infants: assessment and relevance to health and disease. Arch Dis Child Fetal Neonatal Ed **98**(4): F286-290.

Bhatter, P., Chatterjee, A., D'Souza, D., Tolani, M., and Mistry, N. 2012. Estimating fitness by competition assays between drug susceptible and resistant *Mycobacterium tuberculosis* of predominant lineages in Mumbai, India. PLoS One **7**(3): e33507.

Bhaya, D., Davison, M., and Barrangou, R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annual Review of Genetics **45**: 273-297.

Blaiotta, G., Fusco, V., Ercolini, D., Aponte, M., Pepe, O., and Villani, F. 2008. *Lactobacillus* strain diversity based on partial hsp60 gene sequences and design of PCR-restriction fragment length polymorphism assays for species identification and differentiation. Appl Environ Microbiol **74**(1): 208-215.

Bliss, E.A. 1937. Studies upon Minute Hemolytic Streptococci: III. Serological Differentiation. J Bacteriol **33**(6): 625-642.

Bollmann, A., Lewis, K., and Epstein, S.S. 2007. Incubation of environmental samples in a diffusion chamber increases the diversity of recovered isolates. Appl Environ Microbiol **73**(20): 6386-6390.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151**(Pt 8): 2551-2561.

Bosch, T., et al. 2013. High Resolution Typing by Whole Genome Mapping Enables Discrimination of LA-MRSA (CC398) Strains and Identification of Transmission Events. PLoS One **8**(6): e66493.

Bourgogne, A., et al. 2008. Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. Genome Biol **9**(7): R110.

Britschgi, T.B., and Giovannoni, S.J. 1991. Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. Appl Environ Microbiol **57**(6): 1707-1713.

Brodt, A., Lurie-Weinberger, M.N., and Gophna, U. 2011. CRISPR loci reveal networks of gene exchange in archaea. Biol Direct **6**(1): 65.

Broom, L.J., Miller, H.M., Kerr, K.G., and Knapp, J.S. 2006. Effects of zinc oxide and *Enterococcus faecium* SF68 dietary supplementation on the performance, intestinal microbiota and immune status of weaned piglets. Res Vet Sci **80**: 45-54.

Brosius, J., Palmer, M.L., Kennedy, P.J., and Noller, H.F. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc Natl Acad Sci U S A **75**(10): 4801-4805.

Brousseau, R., Hill, J.E., Prefontaine, G., Goh, S.H., Harel, J., and Hemmingsen, S.M. 2001. *Streptococcus suis* serotypes characterized by analysis of chaperonin 60 gene sequences. Appl Environ Microbiol **67**(10): 4828-4833.

Brudey, K., et al. 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol **6**: 23.

Bryant, J.A., Lamanna, C., Morlon, H., Kerkhoff, A.J., Enquist, B.J., and Green, J.L. 2008. Colloquium paper: microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. Proc Natl Acad Sci U S A **105 Suppl 1**: 11505-11511.

Burkler, J., and Solioz, M. 1982. The ATP-dependent Ca2+-pumping system of *Streptococcus faecium*. Ann N Y Acad Sci **402**: 422-432.

Burley, K.M., and Sedgley, C.M. 2012. CRISPR-Cas, a prokaryotic adaptive immune system, in endodontic, oral, and multidrug-resistant hospital-acquired Enterococcus faecalis. Journal of Endodontics **38**(11): 1511-1515.

Butine, T.J., and Leedle, J.A. 1989. Enumeration of selected anaerobic bacterial groups in cecal and colonic contents of growing-finishing pigs. Appl Environ Microbiol **55**(5): 1112-1116.

Cadillo-Quiroz, H., et al. 2012. Patterns of gene flow define species of thermophilic Archaea. PLoS Biol **10**(2): e1001265.

Canalejo, E., Ballesteros, R., Cabezudo, J., Garcia-Arata, M.I., and Moreno, J. 2008. Bacteraemic spondylodiscitis caused by *Enterococcus hirae*. Eur J Clin Microbiol Infect Dis **27**(7): 613-615.

Cangelosi, G.A., et al. 2004. Evaluation of a high-throughput repetitive-sequence-based PCR system for DNA fingerprinting of *Mycobacterium tuberculosis* and *Mycobacterium avium* complex strains. J Clin Microbiol **42**(6): 2685-2693.

Cariveau, D.P., Elijah Powell, J., Koch, H., Winfree, R., and Moran, N.A. 2014. Variation in gut microbial communities and its association with pathogen infection in wild bumble bees (*Bombus*). ISME J **8**(12): 2369-2379.

Carvalho Mda, G., et al. 2006. *Enterococcus caccae* sp. nov., isolated from human stools. Int J Syst Evol Microbiol **56**(Pt 7): 1505-1508.

Carvalho Mda, G., et al. 2008. Designation of the provisional new enterococcus species CDC PNS-E2 as *Enterococcus sanguinicola* sp. nov., isolated from human blood, and identification of a strain previously named *Enterococcus* CDC PNS-E1 as *Enterococcus italicus* Fortina, Ricci, Mora, and Manachini 2004. J Clin Microbiol **46**(10): 3473-3476.

Casamayor, E.O., Pedros-Alio, C., Muyzer, G., and Amann, R. 2002. Microheterogeneity in 16S ribosomal DNA-defined bacterial populations from a stratified planktonic environment is related to temporal changes and to ecological adaptations. Appl Environ Microbiol **68**(4): 1706-1714.

Case, R.J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W.F., and Kjelleberg, S. 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol **73**(1): 278-288.

Cases, I., Velazquez, F., and de Lorenzo, V. 2007. The ancestral role of the phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS) as exposed by comparative genomics. Res Microbiol **158**(8-9): 666-670.

Chaban, B., Albert, A., Links, M.G., Gardy, J., Tang, P., and Hill, J.E. 2013. Characterization of the upper respiratory tract microbiomes of patients with pandemic H1N1 influenza. PLoS One **8**(7): e69559.

Chaban, B., and Hill, J.E. 2012. A 'universal' type II chaperonin PCR detection system for the investigation of Archaea in complex microbial communities. ISME J **6**(2): 430-439.

Chaban, B., Links, M.G., and Hill, J.E. 2012. A molecular enrichment strategy based on cpn60 for detection of epsilon-proteobacteria in the dog fecal microbiome. Microb Ecol **63**(2): 348-357.

Chaban, B., Musil, K.M., Himsworth, C.G., and Hill, J.E. 2009. Development of cpn60-based real-time quantitative PCR assays for the detection of 14 *Campylobacter* species and application to screening of canine fecal samples. Appl Environ Microbiol **75**(10): 3055-3061.

Chaban, B., Ngeleka, M., and Hill, J.E. 2010. Detection and quantification of 14 *Campylobacter* species in pet dogs reveals an increase in species richness in feces of diarrheic animals. BMC Microbiol **10**: 73.

Chain, P.S., et al. 2009. Genomics. Genome project standards in a new era of sequencing. Science **326**(5950): 236-237.

Chan, T.S., et al. 2012. *Enterococcus hirae*-related acute pyelonephritis and cholangitis with bacteremia: an unusual infection in humans. Kaohsiung Journal of Medical Science **28**(2): 111-114.

Coburn, P.S., Baghdayan, A.S., Dolan, G.T., and Shankar, N. 2007. Horizontal transfer of virulence genes encoded on the *Enterococcus faecalis* pathogenicity island. Mol Microbiol **63**(2): 530-544.

Cohan, F.M. 2002a. Sexual isolation and speciation in bacteria. Genetica **116**(2-3): 359-370.

Cohan, F.M. 2002b. What are bacterial species? Annu Rev Microbiol **56**: 457-487.

Cohan, F.M. 2006. Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. Philos Trans R Soc Lond B Biol Sci **361**(1475): 1985-1996.

Cohan, F.M. 2009. Tracking bacterial responses to global warming with an ecotype-based systematics. Clin Microbiol Infect **15 Suppl 1**: 54-59.

Cohan, F.M., and Perry, E.B. 2007. A systematics for discovering the fundamental units of bacterial diversity. Curr Biol **17**(10): R373-386.

Cole, J.R., et al. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. Nucleic Acids Res **35**(Database issue): D169-172.

Collins, M.D., Ash, C., Farrow, J.A., Wallbanks, S., and Williams, A.M. 1989a. 16S ribosomal ribonucleic acid sequence analyses of lactococci and related taxa. Description of *Vagococcus fluvialis* gen. nov., sp. nov. J Appl Bacteriol **67**(4): 453-460.

Collins, M.D., Facklam, R.R., Farrow, J.A., and Williamson, R. 1989b. *Enterococcus raffinosus* sp. nov., *Enterococcus solitarius* sp. nov. and *Enterococcus pseudoavium* sp. nov. FEMS Microbiol Lett **48**(3): 283-288.

Collins, M.D., Farrow, J.A.E., and Jones, D. 1986. *Enterococcus mundtii* sp. nov. Int. J. Syst. Bacteriol(36): 8-12.

Collins, M.D., Jones, D., Farrow, J.A.E., Kilpper-, B.R., and Schleifer, K.H. 1984. *Enterococcus avium* nom. rev., comb. nov.; *E. casseliflavus* nom. rev., comb. nov.; *E.*

durans nom. rev., comb. nov.; *E. gallinarum* comb. nov.; and *E. malodoratus* sp. nov. Int. J. Syst. Bacteriol **34**: 220-223.

Collins, M.D., Rodrigues, U.M., Pigott, N.E., and Facklam, R.R. 1991. *Enterococcus dispar* sp. nov. a new Enterococcus species from human sources. Lett Appl Microbiol **12**(3): 95-98.

Comas, I., Gonzalez-Candelas, F., and Zuniga, M. 2008. Unraveling the evolutionary history of the phosphoryl-transfer chain of the phosphoenolpyruvate:phosphotransferase system through phylogenetic analyses and genome context. BMC Evol Biol **8**: 147.

Connon, S.A., and Giovannoni, S.J. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. Appl Environ Microbiol **68**(8): 3878-3885.

Connor, N., et al. 2010. Ecology of speciation in the genus *Bacillus*. Appl Environ Microbiol **76**(5): 1349-1358.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. 2009. Bacterial community variation in human body habitats across space and time. Science **326**(5960): 1694-1697.

Cotta, M.A., Whitehead, T.R., Falsen, E., Moore, E., and Lawson, P.A. 2013. Two novel species *Enterococcus lemanii* sp. nov. and *Enterococcus eurekensis* sp. nov., isolated from a swine-manure storage pit. Antonie Van Leeuwenhoek **103**(6): 1409-1418.

Crawford, J.T. 2003. Genotyping in contact investigations: a CDC perspective. International Journal of Tuberculosis and Lung Disease **7**(12 Suppl 3): S453-457.

Crow, V.L., Davey, G.P., Pearce, L.E., and Thomas, T.D. 1983. Plasmid linkage of the D-tagatose 6-phosphate pathway in *Streptococcus lactis*: effect on lactose and galactose metabolism. J Bacteriol **153**(1): 76-83.

Dahllof, I., Baillie, H., and Kjelleberg, S. 2000. rpoB-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. Appl Environ Microbiol **66**(8): 3376-3380.

Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K., and Semenova, E. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nature Communications **3**: 945.

de Graef, E.M., et al. 2003. Description of *Enterococcus canis* sp. nov. from dogs and reclassification of *Enterococcus porcinus* Teixeira et al. 2001 as a junior synonym of *Enterococcus villorum* Vancanneyt et al 2001. Int J Syst Evol Microbiol **53**(Pt 4): 1069-1074.

de Vaux, A., Laguerre, G., Divies, C., and Prevost, H. 1998. *Enterococcus asini* sp. nov. isolated from the caecum of donkeys (Equus asinus). Int J Syst Bacteriol **48 Pt 2**: 383-387.

de Vos, W.M., and Vaughan, E.E. 1994. Genetics of lactose utilization in lactic acid bacteria. FEMS Microbiol Rev **15**(2-3): 217-237.

De Vuyst, L., et al. 2008. Validation of the (GTG)(5)-rep-PCR fingerprinting technique for rapid classification and identification of acetic acid bacteria, with a focus on isolates from Ghanaian fermented cocoa beans. Int J Food Microbiol **125**(1): 79-90.

Delannoy, S., Beutin, L., Burgos, Y., and Fach, P. 2012a. Specific detection of enteroaggregative hemorrhagic *Escherichia coli* O104:H4 strains using the CRISPR locus as target for a diagnostic real-time PCR. J Clin Microbiol **50**(11): 3485-3492.

Delannoy, S., Beutin, L., and Fach, P. 2012b. Use of CRISPR sequence polymorphisms for specific detection of enterohemorrhagic *E. coli* (EHEC) strains of serotypes O26:H11, O45:H2, O103:H2, O111:H8, O121:H19, O145:H28 and O157:H7 by real time PCR. J Clin Microbiol **Epub ahead of print**: JCM.02097-02012.

Denef, V.J., et al. 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. Proc Natl Acad Sci U S A **107**(6): 2383-2390.

Desai, A.R., et al. 2012. Effects of plant-based diets on the distal gut microbiome of rainbow trout (*Oncorhynchus mykiss*). Aquaculture **350-353**: 134-142.

Desai, A.R., Musil, K.M., Carr, A.P., and Hill, J.E. 2009. Characterization and quantification of feline fecal microbiota using cpn60 sequence-based methods and investigation of animal-to-animal variation in microbial population structure. Vet Microbiol **137**(1-2): 120-128.

Devriese, L.A., Ceyssens, K., Rodrigues, U.M., and Collins, M.D. 1990. *Enterococcus columbae*, a species from pigeon intestines. FEMS Microbiol Lett **59**(3): 247-251.

Devriese, L.A., Cruz Colque, J.I., De Herdt, P., and Haesebrouck, F. 1992a. Identification and composition of the tonsillar and anal enterococcal and streptococcal flora of dogs and cats. J Appl Bacteriol **73**(5): 421-425.

Devriese, L.A., Cruz Colque, J.I., Haesebrouck, F., Desmidt, M., Uyttebroek, E., and Ducatelle, R. 1992b. *Enterococcus hirae* in septicaemia of psittacine birds. Vet Rec **130**(25): 558-559.

Devriese, L.A., and Haesebrouck, F. 1991. *Enterococcus hirae* in different animal species. Vet Rec **129**(17): 391-392.

Devriese, L.A., Hommez, J., Pot, B., and Haesebrouck, F. 1994. Identification and composition of the streptococcal and enterococcal flora of tonsils, intestines and faeces of pigs. J Appl Bacteriol **77**(1): 31-36.

Devriese, L.A., Hommez, J., Wijfels, R., and Haesebrouck, F. 1991. Composition of the enterococcal and streptococcal intestinal flora of poultry. J Appl Bacteriol **71**(1): 46-50.

Devriese, L.A., Laurier, L., De Herdt, P., and Haesebrouck, F. 1992c. Enterococcal and streptococcal species isolated from faeces of calves, young cattle and dairy cows. J Appl Bacteriol **72**(1): 29-31.

Devriese, L.A., Pot, B., and Collins, M.D. 1993. Phenotypic identification of the genus *Enterococcus* and differentiation of phylogenetically distinct enterococcal species and species groups. J Appl Bacteriol **75**(5): 399-408.

Devriese, L.A., Pot, B., Van Damme, L., Kersters, K., and Haesebrouck, F. 1995. Identification of *Enterococcus* species isolated from foods of animal origin. Int J Food Microbiol **26**(2): 187-197.

Doolittle, W.F., and Zhaxybayeva, O. 2009. On the origin of prokaryotic species. Genome Res **19**(5): 744-756.

Dumonceaux, T.J., Green, M., Hammond, C., Perez, E., and Olivier, C. 2014. Molecular diagnostic tools for detection and differentiation of phytoplasmas based on chaperonin-60 reveal differences in host plant infection patterns. PLoS One **9**(12): e116039.

Dumonceaux, T.J., Hill, J.E., Hemmingsen, S.M., and Van Kessel, A.G. 2006a. Characterization of intestinal microbiota and response to dietary virginiamycin supplementation in the broiler chicken. Appl Environ Microbiol **72**(4): 2815-2823.

Dumonceaux, T.J., Hill, J.E., Pelletier, C., Paice, M.G., Van Kessel, A.G., and Hemmingsen, S.M. 2006b. Molecular characterization of microbial communities in Canadian pulp and paper activated sludge and quantification of a novel *Thiothrix eikelboomii*-like bulking filament. Canadian Journal of Microbiology **52**(5): 494-500.

Dumonceaux, T.J., et al. 2009. Multiplex detection of bacteria associated with normal microbiota and with bacterial vaginosis in vaginal swabs by use of oligonucleotide-coupled fluorescent microspheres. J Clin Microbiol **47**(12): 4067-4077.

Dumonceaux, T.J., Town, J.R., Hill, J.E., Chaban, B.L., and Hemmingsen, S.M. 2011. Multiplex detection of bacteria in complex clinical and environmental samples using oligonucleotide-coupled fluorescent microspheres. The Journal of Visualized Experiments(56).

Dunbar, J., Ticknor, L.O., and Kuske, C.R. 2000. Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. Appl Environ Microbiol **66**(7): 2943-2950.

Dykhuizen, D.E. 1998. Santa Rosalia revisited: why are there so many species of bacteria? Antonie Van Leeuwenhoek **73**(1): 25-33.

Elliott, J.A., Collins, M.D., Pigott, N.E., and Facklam, R.R. 1991. Differentiation of *Lactococcus lactis* and *Lactococcus garvieae* from humans by comparison of whole-cell protein patterns. J Clin Microbiol **29**(12): 2731-2734.

Ennahar, S., and Cai, Y. 2005. Biochemical and genetic evidence for the transfer of *Enterococcus solitarius* Collins et al. 1989 to the genus *Tetragenococcus* as *Tetragenococcus solitarius* comb. nov. Int J Syst Evol Microbiol **55**(Pt 2): 589-592.

Facklam, R., and Elliott, J.A. 1995. Identification, classification, and clinical relevance of catalase-negative, gram-positive cocci, excluding the streptococci and enterococci. Clin Microbiol Rev **8**(4): 479-495.

Facklam, R.R., and Collins, M.D. 1989. Identification of Enterococcus species isolated from human infections by a conventional test scheme. J Clin Microbiol **27**(4): 731-734.

Facklam, R.R., Collins, M.D., and Teixeira, L.M. 2002. History, Taxonomy, Biochemical Characteristics, and Antibiotic Susceptibility Testing of Enterococci (Pages 1-54) in The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance. ASM Press.

Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. 2011. Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science **333**(6038): 101-104.

Farrow, J.A.E., and Collins, M.D. 1985. *Enterococcus hirae*, a new species that includes amino acid assay strain NCDO 1258 and strains causing growth depression in young chickens. Int. J. Syst. Bacteriol **35**: 73-75.

Favier, C.F., Vaughan, E.E., De Vos, W.M., and Akkermans, A.D. 2002. Molecular monitoring of succession of bacterial communities in human neonates. Appl Environ Microbiol **68**(1): 219-226.

Federle, T.W., McKinley, V.L., and Vestal, J.R. 1982. Physical determinants of microbial colonization and decomposition of plant litter in an Arctic Lake. Microb Ecol **8**(2): 127-138.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Ferrari, B.C., Binnerup, S.J., and Gillings, M. 2005. Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. Appl Environ Microbiol **71**(12): 8714-8720.

Field, K.G., et al. 1997. Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. Appl Environ Microbiol **63**(1): 63-70.

Fierer, N., et al. 2011. Microbes do not follow the elevational diversity patterns of plants and animals. Ecology **92**(4): 797-804.

Fierer, N., Nemergut, D., Knight, R., and Craine, J.M. 2010. Changes through time: integrating microorganisms into the study of succession. Res Microbiol **161**(8): 635-642.

Fisher, M.M., and Triplett, E.W. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl Environ Microbiol **65**(10): 4630-4636.

Fortina, M.G., Ricci, G., Mora, D., and Manachini, P.L. 2004. Molecular analysis of artisanal Italian cheeses reveals *Enterococcus italicus* sp. nov. Int J Syst Evol Microbiol **54**(Pt 5): 1717-1721.

Fox, G.E., Wisotzkey, J.D., and Jurtshuk, P., Jr. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol **42**(1): 166-170.

Francl, A.L., Hoeflinger, J.L., and Miller, M.J. 2012. Identification of lactose phosphotransferase systems in *Lactobacillus gasseri* ATCC 33323 required for lactose utilization. Microbiology **158**(Pt 4): 944-952.

Francl, A.L., Thongaram, T., and Miller, M.J. 2010. The PTS transporters of *Lactobacillus gasseri* ATCC 33323. BMC Microbiol **10**: 77.

Franz, C.M., Stiles, M.E., Schleifer, K.H., and Holzapfel, W.H. 2003. Enterococci in foods--a conundrum for food safety. Int J Food Microbiol **88**(2-3): 105-122.

Franz, C.M., van Belkum, M.J., Holzapfel, W.H., Abriouel, H., and Galvez, A. 2007. Diversity of enterococcal bacteriocins and their grouping in a new classification scheme. FEMS Microbiol Rev **31**(3): 293-310.

Freter, R., and O'Brien, P.C. 1981. Role of chemotaxis in the association of motile bacteria with intestinal mucosa: chemotactic responses of *Vibrio cholerae* and description of motile nonchemotactic mutants. Infect Immun **34**(1): 215-221.

Frolkova, P., Ghosh, A., Svec, P., Zurek, L., and Literak, I. 2012. Use of the manganese-dependent superoxide dismutase gene sodA for rapid identification of recently described enterococcal species. Folia Microbiol (Praha) **57**(5): 439-442.

Frolkova, P., et al. 2013. *Enterococcus alcedinis* sp. nov., isolated from common kingfisher (Alcedo atthis). Int J Syst Evol Microbiol **63**(Pt 8): 3069-3074.

Gaechter, T., Wunderlin, C., Schmidheini, T., and Solioz, M. 2012. Genome sequence of *Enterococcus hirae* (*Streptococcus faecalis*) ATCC 9790, a model organism for the study of ion transport, bioenergetics, and copper homeostasis. J Bacteriol **194**(18): 5126-5127.

Galloway-Pena, J., Roh, J.H., Latorre, M., Qin, X., and Murray, B.E. 2012. Genomic and SNP analyses demonstrate a distant separation of the hospital and community-associated clades of *Enterococcus faecium*. PLoS One **7**(1): e30187.

Gevers, D., Huys, G., and Swings, J. 2001. Applicability of rep-PCR fingerprinting for identification of *Lactobacillus* species. FEMS Microbiol Lett **205**(1): 31-36.

Gevers, D., Pop, M., Schloss, P.D., and Huttenhower, C. 2012. Bioinformatics for the Human Microbiome Project. PLoS Comput Biol **8**(11): e1002779.

Gilad, J., Borer, A., Riesenberg, K., Peled, N., Shnaider, A., and Schlaeffer, F. 1998. *Enterococcus hirae* septicemia in a patient with end-stage renal disease undergoing hemodialysis. Eur J Clin Microbiol Infect Dis **17**(8): 576-577.

Gilmore, M.S. 2002. The Enterococci: Pathogenesis, Molecular Biology, and Antibiotic Resistance. ASM Press.

Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature **345**(6270): 60-63.

Glazunova, O.O., Raoult, D., and Roux, V. 2009. Partial sequence comparison of the rpoB, sodA, groEL and gyrB genes within the genus Streptococcus. Int J Syst Evol Microbiol **59**(Pt 9): 2317-2322.

Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. 2002. Prokaryotic evolution in light of gene transfer. Mol Biol Evol **19**(12): 2226-2238.

Goh, S.H., et al. 1998. *Streptococcus iniae*, a human and animal pathogen: specific identification by the chaperonin 60 gene identification method. J Clin Microbiol **36**(7): 2164-2166.

Goh, S.H., et al. 2000. Identification of *Enterococcus* species and phenotypically similar *Lactococcus* and *Vagococcus* species by reverse checkerboard hybridization to chaperonin 60 gene sequences. J Clin Microbiol **38**(11): 3953-3959.

Goh, S.H., Potter, S., Wood, J.O., Hemmingsen, S.M., Reynolds, R.P., and Chow, A.W. 1996. HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. J Clin Microbiol **34**(4): 818-823.

Goh, S.H., et al. 1997. Identification of *Staphylococcus* species and subspecies by the chaperonin 60 gene identification method and reverse checkerboard hybridization. J Clin Microbiol **35**(12): 3116-3121.

Gordillo, M.E., Singh, K.V., and Murray, B.E. 1993. Comparison of ribotyping and pulsed-field gel electrophoresis for subspecies differentiation of strains of *Enterococcus faecalis*. J Clin Microbiol **31**(6): 1570-1574.

Graudal, H. 1957. The classification of motile streptococci within the enterococcus group. Acta Pathol Microbiol Scand **41**: 403–410.

Grissa, I., Vergnaud, G., and Pourcel, C. 2007a. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics **8**: 172.

Grissa, I., Vergnaud, G., and Pourcel, C. 2007b. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res **35**: W52-W57.

Gundi, V.A., Dijkshoorn, L., Burignat, S., Raoult, D., and La Scola, B. 2009. Validation of partial rpoB gene sequence analysis for the identification of clinically important and emerging Acinetobacter species. Microbiology **155**(Pt 7): 2333-2341.

Hammerum, A.M. 2012. Enterococci of animal origin and their significance for public health. Clin Microbiol Infect **18**(7): 619-625.

Harold, F.M., Harold, R.L., and Abrams, A. 1965. A mutant of *Streptococcus faecalis* defective in phosphate uptake. J Biol Chem **240**: 3145-3153.

Harris, J.K., Kelley, S.T., and Pace, N.R. 2004. New perspective on uncultured bacterial phylogenetic division OP11. Appl Environ Microbiol **70**(2): 845-849.

Harrison, A.P., Jr. 1978. Microbial succession and mineral leaching in an artificial coal spoil. Appl Environ Microbiol **36**(6): 861-869.

Hatfield, D.L., and Gladyshev, V.N. 2002. How selenium has altered our understanding of the genetic code. Mol Cell Biol **22**(11): 3565-3576.

Hauck, Y., et al. 2012. Diversity of *Acinetobacter baumannii* in four French military hospitals, as assessed by multiple locus variable number of tandem repeats analysis. PLoS One **7**(9): e44597.

Heefner, D.L., and Harold, F.M. 1982. ATP-driven sodium pump in *Streptococcus faecalis*. Proc Natl Acad Sci U S A **79**(9): 2798-2802.

Heinritz, S.N., Mosenthin, R., and Weiss, E. 2013. Use of pigs as a potential model for research into dietary modulation of the human gut microbiota. Nutr Res Rev **26**(2): 191-209.

Hemmingsen, S.M., et al. 1988. Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature **333**(6171): 330-334.

Hennequin, C., et al. 2001. GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. Microbiology **147**(Pt 1): 87-96.

Hew, C.M., Korakli, M., and Vogel, R.F. 2007. Expression of virulence-related genes by *Enterococcus faecalis* in response to different environments. Syst Appl Microbiol **30**(4): 257-267.

Hill, J.E., Fernando, W.M., Zello, G.A., Tyler, R.T., Dahl, W.J., and Van Kessel, A.G. 2010. Improvement of the representation of bifidobacteria in fecal microbiota metagenomic libraries by application of the cpn60 universal primer cocktail. Appl Environ Microbiol **76**(13): 4550-4552.

Hill, J.E., et al. 2005. Comparison of ileum microflora of pigs fed corn-, wheat-, or barley-based diets by chaperonin-60 sequencing and quantitative PCR. Appl Environ Microbiol **71**(2): 867-875.

Hill, J.E., et al. 2006a. Identification of *Campylobacter* spp. and discrimination from *Helicobacter* and *Arcobacter* spp. by direct sequencing of PCR-amplified cpn60 sequences and comparison to cpnDB, a chaperonin reference sequence database. J Med Microbiol **55**(Pt 4): 393-399.

Hill, J.E., Penny, S.L., Crowell, K.G., Goh, S.H., and Hemmingsen, S.M. 2004. cpnDB: a chaperonin sequence database. Genome Res **14**(8): 1669-1675.

Hill, J.E., et al. 2002. Extensive profiling of a complex microbial community by high-throughput sequencing. Appl Environ Microbiol **68**(6): 3055-3066.

Hill, J.E., Town, J.R., and Hemmingsen, S.M. 2006b. Improved template representation in cpn60 polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. Environ Microbiol **8**(4): 741-746.

Hohwy, J., Reinholdt, J., and Kilian, M. 2001. Population dynamics of Streptococcus mitis in its natural habitat. Infect Immun **69**(10): 6055-6063.

Homan, W.L., et al. 2002. Multilocus sequence typing scheme for *Enterococcus faecium*. J Clin Microbiol **40**(6): 1963-1971.

Horvath, P., Coute-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., and Barrangou, R. 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. Int J Food Microbiol **131**(1): 62-70.

Hugenholtz, P., Goebel, B.M., and Pace, N.R. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol **180**(18): 4765-4774.

Hugenholtz, P., and Pace, N.R. 1996. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol **14**(6): 190-197.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., and Polz, M.F. 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science **320**(5879): 1081-1085.

Inoue, R., Tsukahara, T., Nakanishi, N., and Ushida, K. 2005. Development of the intestinal microbiota in the piglet. J Gen Appl Microbiol **51**(4): 257-265.

Isaacson, R., and Kim, H.B. 2012. The intestinal microbiome of the pig. Anim Health Res Rev **13**(1): 100-109.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. J Bacteriol **169**(12): 5429-5433.

Jackson, C.R., Fedorka-Cray, P.J., and Barrett, J.B. 2004. Use of a genus- and species-specific multiplex PCR for identification of enterococci. J Clin Microbiol **42**(8): 3558-3565.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337**(6096): 816-821.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I., and Koonin, E.V. 2002. Microevolutionary genomics of bacteria. Theor Popul Biol **61**(4): 435-447.

Kandler, O. 1983. Carbohydrate metabolism in lactic acid bacteria. Antonie Van Leeuwenhoek **49**(3): 209-224.

Karden-Lilja, M., Vuopio, J., Koskela, M., Tissari, P., and Salmenlinna, S. 2013. Molecular typing of vancomycin-resistant *Enterococcus faecium* with an automated repetitive sequence-based PCR microbial typing system compared with pulsed-field gel electrophoresis and multilocus sequence typing. Scand J Infect Dis **45**(5): 350-356.

Katyal, I., Chaban, B., Ng, B., and Hill, J.E. 2013. CRISPRs of *Enterococcus faecalis* and *E. hirae* isolates from pig feces have species-specific repeats but share some common spacer sequences. Microb Ecol **66**(1): 182-188.

Ke, D., et al. 1999. Development of a PCR assay for rapid detection of enterococci. J Clin Microbiol **37**(11): 3497-3503.

Kelly, D., and Mulder, I.E. 2011. Gut microbiota and GI health and disease. J Pediatr Gastroenterol Nutr **53 Suppl 2**: S32-34.

Kettler, G.C., et al. 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet **3**(12): e231.

Kim, H.B., et al. 2011. Longitudinal investigation of the age-related bacterial diversity in the feces of commercial pigs. Vet Microbiol **153**(1-2): 124-133.

Kim, J.Y., et al. 2013. *Enterococcus diestrammenae* sp. nov., isolated from the gut of Diestrammena coreana. Int J Syst Evol Microbiol **63**(Pt 12): 4540-4545.

Kim, M., Oh, H.S., Park, S.C., and Chun, J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol **64**(Pt 2): 346-351.

Klatt, C.G., et al. 2011. Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J **5**(8): 1262-1278.

Klein, G. 2003. Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. Int J Food Microbiol **88**(2-3): 123-131.

Klima, C.L., et al. 2010. Comparison of repetitive PCR and pulsed-field gel electrophoresis for the genotyping of *Mannheimia haemolytica*. J Microbiol Methods **81**(1): 39-47.

Kobayashi, H., Van Brunt, J., and Harold, F.M. 1978. ATP-linked calcium transport in cells and membrane vesicles of *Streptococcus faecalis*. J Biol Chem **253**(7): 2085-2092.

Koeppel, A., et al. 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. Proc Natl Acad Sci U S A **105**(7): 2504-2509.

Koort, J., Coenye, T., Vandamme, P., Sukura, A., and Bjorkroth, J. 2004. *Enterococcus hermanniensis* sp. nov., from modified-atmosphere-packaged broiler meat and canine tonsils. Int J Syst Evol Microbiol **54**(Pt 5): 1823-1827.

Kopac, S., Wang, Z., Wiedenbeck, J., Sherry, J., Wu, M., and Cohan, F.M. 2014. Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. Appl Environ Microbiol **80**(16): 4842-4853.

Kuczynski, J., et al. 2012. Experimental and analytical tools for studying the human microbiome. Nat Rev Genet **13**(1): 47-58.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol **12**(1): 118-123.

Kunin, V., Sorek, R., and Hugenholtz, P. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol **8**(4): R61.

Kusuda, R., Kawai, K., Salati, F., Banner, C.R., and Fryer, J.L. 1991. *Enterococcus seriolicida* sp. nov., a fish pathogen. Int J Syst Bacteriol **41**(3): 406-409.

Law-Brown, J., and Meyers, P.R. 2003. *Enterococcus phoeniculicola* sp. nov., a novel member of the enterococci isolated from the uropygial gland of the Red-billed Woodhoopoe, *Phoeniculus purpureus*. Int J Syst Evol Microbiol **53**(Pt 3): 683-685.

Leavis, H.L., Willems, R.J., van Wamel, W.J., Schuren, F.H., Caspers, M.P., and Bonten, M.J. 2007. Insertion sequence-driven diversification creates a globally dispersed emerging multiresistant subspecies of *E. faecium*. PLoS Pathog **3**(1): e7.

Lebreton, F., et al. 2013. Emergence of epidemic multidrug-resistant *Enterococcus faecium* from animal and commensal strains. MBio **4**(4).

Lebreton, F., Willems, R.J.L., and Gilmore, M.S. 2014. *Enterococcus* Diversity, Origins in Nature, and Gut Colonization. *In* Enterococci: From Commensals to Leading Causes of Drug Resistant Infection. *Edited by* M. S. Gilmore, D. B. Clewell, Y. Ike, and N. Shankar, Boston.

Lee, D.H., Zo, Y.G., and Kim, S.J. 1996. Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism. Appl Environ Microbiol **62**(9): 3112-3120.

Lepage, E., et al. 2006. Comparative genomic hybridization analysis of *Enterococcus faecalis*: identification of genes absent from food strains. J Bacteriol **188**(19): 6858-6868.

Leser, T.D., Amenuvor, J.Z., Jensen, T.K., Lindecrona, R.H., Boye, M., and Moller, K. 2002. Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. Appl Environ Microbiol **68**(2): 673-690.

Li, C.Y., Tian, F., Zhao, Y.D., and Gu, C.T. 2014. *Enterococcus xiangfangensis* sp. nov., isolated from Chinese pickle. Int J Syst Evol Microbiol **64**(Pt 3): 1012-1017.

Lindenstrauss, A.G., Pavlovic, M., Bringmann, A., Behr, J., Ehrmann, M.A., and Vogel, R.F. 2011. Comparison of genotypic and phenotypic cluster analyses of virulence determinants and possible role of CRISPR elements towards their incidence in *Enterococcus faecalis* and *Enterococcus faecium*. Syst Appl Microbiol **34**(8): 553-560.

Ling, L.L., et al. 2015. A new antibiotic kills pathogens without detectable resistance. Nature **517**(7535): 455-459.

Links, M.G., Dumonceaux, T.J., Hemmingsen, S.M., and Hill, J.E. 2012. The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PLoS One **7**(11): e49755.

Lu, Z.H., Dameron, C.T., and Solioz, M. 2003. The *Enterococcus hirae* paradigm of copper homeostasis: copper chaperone turnover, interactions, and transactions. Biometals **16**(1): 137-143.

Lucena-Padros, H., Gonzalez, J.M., Caballero-Guerrero, B., Ruiz-Barba, J.L., and Maldonado-Barragan, A. 2014. *Enterococcus olivae* sp. nov., isolated from Spanish-style green-olive fermentations. Int J Syst Evol Microbiol **64**(Pt 8): 2534-2539.

Luckey, T.D. 1972. Introduction to intestinal microecology. Am J Clin Nutr **25**(12): 1292-1294.

Maccallum, W.G., and Hastings, T.W. 1899. A Case of Acute Endocarditis Caused by *Micrococcus Zymogenes* (Nov. Spec.), with a Description of the Microorganism. J Exp Med **4**(5-6): 521-534.

Macha, M., Taras, D., Vahjen, W., Arini, A., and Simon, O. 2004. Specific enumeration of the probiotic strain *Enterococcus faecium* NCIMB 10415 in the intestinal tract and in faeces of piglets and sows. Arch Anim Nutr **58**(6): 443-452.

Macho, A.P., Zumaquero, A., Ortiz-Martin, I., and Beuzon, C.R. 2007. Competitive index in mixed infections: a sensitive and accurate assay for the genetic analysis of *Pseudomonas syringae*-plant interactions. Mol Plant Pathol **8**(4): 437-450.

Mafamane, H., et al. 2011. Studies on the effect of an *Enterococcus faecium* probiotic on T cell populations in peripheral blood and intestinal epithelium and on the susceptibility to *Salmonella* during a challenge infection with *Salmonella Typhimurium* in piglets. Arch Anim Nutr **65**(6): 415-430.

Magni, C., et al. 2012. Draft genome sequence of *Enterococcus mundtii* CRL1656. J Bacteriol **194**(2): 550.

Makarova, K.S., et al. 2011. Evolution and classification of the CRISPR-Cas systems. Nature Reviews in Microbiology **9**(6): 467-477.

Malathum, K., Singh, K.V., Weinstock, G.M., and Murray, B.E. 1998. Repetitive sequence-based PCR versus pulsed-field gel electrophoresis for typing of *Enterococcus faecalis* at the subspecies level. J Clin Microbiol **36**(1): 211-215.

Manero, A., and Blanch, A.R. 1999. Identification of *Enterococcus* spp. with a biochemical key. Appl Environ Microbiol **65**(10): 4425-4430.

Mannu, L., and Paba, A. 2002. Genetic diversity of *lactococci* and *enterococci* isolated from home-made Pecorino Sardo ewes' milk cheese. J Appl Microbiol **92**(1): 55-62.

Mannu, L., Paba, A., Pes, M., Floris, R., Scintu, M.F., and Morelli, L. 1999. Strain typing among *enterococci* isolated from home-made Pecorino Sardo cheese. FEMS Microbiol Lett **170**(1): 25-30.

Marcellino, S.N., and Benson, D.R. 1992. Scanning electron and light microscopic study of microbial succession on bethlehem st. Nectaire cheese. Appl Environ Microbiol **58**(11): 3448-3454.

Martin, J.D., and Mundt, J.O. 1972. *Enterococci* in insects. Appl Microbiol **24**(4): 575-580.

Martin-Platero, A.M., Valdivia, E., Maqueda, M., and Martinez-Bueno, M. 2007. Fast, convenient, and economical method for isolating genomic DNA from lactic acid bacteria using a modification of the protein "salting-out" procedure. Anal Biochem **366**(1): 102-104.

Martinez-Murcia, A.J., and Collins, M.D. 1991. *Enterococcus sulfureus*, a new yellow-pigmented Enterococcus species. FEMS Microbiol Lett **64**(1): 69-74.

Masco, L., Huys, G., Gevers, D., Verbrugghen, L., and Swings, J. 2003. Identification of *Bifidobacterium* species using rep-PCR fingerprinting. Syst Appl Microbiol **26**(4): 557-563.

Mazmanian, S.K., Liu, C.H., Tzianabos, A.O., and Kasper, D.L. 2005. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell **122**(1): 107-118.

McBride, S.M., Fischetti, V.A., Leblanc, D.J., Moellering, R.C., Jr., and Gilmore, M.S. 2007. Genetic diversity among *Enterococcus faecalis*. PLoS One **2**(7): e582.

McKay, L., Miller, A., 3rd, Sandine, W.E., and Elliker, P.R. 1970. Mechanisms of lactose utilization by lactic acid streptococci: enzymatic and genetic analyses. J Bacteriol **102**(3): 804-809.

Minana-Galbis, D., Urbizu-Serrano, A., Farfan, M., Fuste, M.C., and Loren, J.G. 2009. Phylogenetic analysis and identification of *Aeromonas* species based on sequencing of the cpn60 universal target. Int J Syst Evol Microbiol **59**(Pt 8): 1976-1983.

Miyuki Kawano, K.I.a.K. 1999. Two major potassium uptake systems, KtrI and KtrII, in *Enterococcus hirae*. FEMS Microbiol Lett **176** (2): 449–453.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology **155**(Pt 3): 733-740.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol **60**(2): 174-182.

Mollet, C., Drancourt, M., and Raoult, D. 1997. rpoB sequence analysis as a novel basis for bacterial identification. Mol Microbiol **26**(5): 1005-1011.

Monk, I.R., Casey, P.G., Cronin, M., Gahan, C.G., and Hill, C. 2008. Development of multiple strain competitive index assays for *Listeria monocytogenes* using pIMC; a new site-specific integrative vector. BMC Microbiol **8**: 96.

Morandi, S., Cremonesi, P., Povolo, M., and Brasca, M. 2012. *Enterococcus lactis* sp. nov., from Italian raw milk cheeses. Int J Syst Evol Microbiol **62**(Pt 8): 1992-1996.

Morato, J., Codony, F., Sanchez, O., Perez, L.M., Garcia, J., and Mas, J. 2014. Key design factors affecting microbial community composition and pathogenic organism removal in horizontal subsurface flow constructed wetlands. Sci Total Environ **481**: 81-89.

Morgan, X.C., and Huttenhower, C. 2014. Meta'omic analytic techniques for studying the intestinal microbiome. Gastroenterology **146**(6): 1437-1448 e1431.

Mundt, J.O. 1961. Occurrence of *Enterococci*: Bud, Blossom, and Soil Studies. Appl Microbiol **9**(6): 541-544.

Murata, T., Yamato, I., Igarashi, K., and Kakinuma, Y. 1996. Intracellular Na+ regulates transcription of the ntp operon encoding a vacuolar-type Na+-translocating ATPase in *Enterococcus hirae*. J Biol Chem **271**(39): 23661-23666.

Murray, B.E. 1990. The life and times of the *Enterococcus*. Clin Microbiol Rev **3**(1): 46-65.

Nagarajan, N., and Pop, M. 2013. Sequence assembly demystified. Nat Rev Genet **14**(3): 157-167.

Naimi, A., Beck, G., Monique, M., Lefebvre, G., and Branlanti, C. 1999. Determination of the nucleotide sequence of the 23S ribosomal RNA and flanking spacers of an *Enterococcus faecium* strain, reveals insertion-deletion events in the ribosomal spacer 1 of enterococci. Syst Appl Microbiol **22**(1): 9-21.

Naser, S.M., et al. 2005a. Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on rpoA and pheS genes. Microbiology **151**(Pt 7): 2141-2150.

Naser, S.M., et al. 2005b. *Enterococcus canintestini* sp. nov., from faecal samples of healthy dogs. Int J Syst Evol Microbiol **55**(Pt 5): 2177-2182.

Naser, S.M., Vancanneyt, M., Hoste, B., Snauwaert, C., Vandemeulebroecke, K., and Swings, J. 2006. Reclassification of *Enterococcus flavescens* Pompei et al. 1992 as a later synonym of *Enterococcus casseliflavus* (ex Vaughan et al. 1979) Collins et al. 1984 and *Enterococcus saccharominimus* Vancanneyt et al. 2004 as a later synonym of *Enterococcus italicus* Fortina et al. 2004. Int J Syst Evol Microbiol **56**(Pt 2): 413-416.

Nayak, B.S., Badgley, B., and Harwood, V.J. 2011. Comparison of genotypic and phylogenetic relationships of environmental *Enterococcus* isolates by BOX-PCR typing and 16S rRNA gene sequencing. Appl Environ Microbiol **77**(14): 5050-5055.

Nemergut, D.R., et al. 2011. Global patterns in the biogeography of bacterial taxa. Environ Microbiol **13**(1): 135-144.

Nes, I.F., Diep, D.B., and Holo, H. 2007. Bacteriocin diversity in *Streptococcus* and *Enterococcus*. J Bacteriol **189**(4): 1189-1198.

Nicholas, K.B., and Nicholas, H.B.J. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. **Distributed by the author.** http://www.psc.edu/biomed/genedoc.

Niemi, R.M., et al. 1993. Presumptive fecal streptococci in environmental samples characterized by one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Appl Environ Microbiol **59**(7): 2190-2196.

Niemi, R.M., et al. 2012. *Enterococcus rivorum* sp. nov., from water of pristine brooks. Int J Syst Evol Microbiol **62**(Pt 9): 2169-2173.

Odermatt, A., Suter, H., Krapf, R., and Solioz, M. 1993. Primary structure of two P-type ATPases involved in copper homeostasis in *Enterococcus hirae*. J Biol Chem **268**(17): 12775-12779.

Orla-Jensen, S. 1919. The lactic acid bacteria. Memoirs of the Academy of the Royal Society of Denmark. Sect Sci Ser. **85:81–197**.

Orr, H.A. 2009. Fitness and its role in evolutionary genetics. Nat Rev Genet **10**(8): 531-539.

Ostojic, M. 2008. Epidemiologic genotyping of methicillin-resistant *Staphylococcus aureus* (MRSA) by pulsed-field gel electrophoresis (PFGE). Bosn J Basic Med Sci **8**(3): 259-265.

Ottesen, E.A., et al. 2013. Pattern and synchrony of gene expression among sympatric marine microbial populations. Proc Natl Acad Sci U S A **110**(6): E488-497.

Ozawa, Y., Courvalin, P., and Gaiimand, M. 2000. Identification of enterococci at the species level by sequencing of the genes for D-alanine:D-alanine ligases. Syst Appl Microbiol **23**(2): 230-237.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., and Van Oudenaarden, A. 2004. Multistability in the lactose utilization network of *Escherichia coli*. Nature **427**(6976): 737-740.

Palleroni, N.J. 1997. Prokaryotic diversity and the importance of culturing. Antonie Van Leeuwenhoek **72**(1): 3-19.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. 2007. Development of the human infant intestinal microbiota. PLoS Biol **5**(7): e177.

Palmer, K.L., et al. 2010. High-quality draft genome sequences of 28 *Enterococcus* sp. isolates. J Bacteriol **192**(9): 2469-2470.

Palmer, K.L., and Gilmore, M.S. 2010. Multidrug-resistant enterococci lack CRISPR-cas. MBio **1**(4).

Palmer, K.L., et al. 2012. Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. gallinarum* and *E. casseliflavus*. MBio **3**(1): e00318-00311.;

Palys, T., Berger, E., Mitrica, I., Nakamura, L.K., and Cohan, F.M. 2000. Protein-coding genes as molecular markers for ecologically distinct populations: the case of two Bacillus species. Int J Syst Evol Microbiol **50 Pt 3**: 1021-1028.

Palys, T., Nakamura, L.K., and Cohan, F.M. 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int J Syst Bacteriol **47**(4): 1145-1156.

Pangallo, D., et al. 2008. Evaluation of different PCR-based approaches for the identification and typing of environmental enterococci. Antonie Van Leeuwenhoek **93**(1-2): 193-203.

Papaparaskevas, J., Tassios, P.T., Kalapothaki, V., Avlami, A., Legakis, N.J., and Vatopoulos, A.C. 2002. Epidemiology of multiresistant *Enterococcus avium* isolates in a Greek tertiary care hospital. Int J Antimicrob Agents **20**(6): 432-437.

Paramel Jayaprakash, T., Schellenberg, J.J., and Hill, J.E. 2012. Resolution and characterization of distinct cpn60-based subgroups of *Gardnerella vaginalis* in the vaginal microbiota. PLoS One **7**(8): e43009.

Parte, A.C. 2014. LPSN--list of prokaryotic names with standing in nomenclature. Nucleic Acids Res **42**(Database issue): D613-616.

Patel, R., et al. 1998. Determination of 16S rRNA sequences of enterococci and application to species identification of nonmotile *Enterococcus gallinarum* isolates. J Clin Microbiol **36**(11): 3399-3407.

Paulsen, I.T., et al. 2003. Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science **299**(5615): 2071-2074.

Pecou, E., Maass, A., Remenik, D., Briche, J., and Gonzalez, M. 2006. A mathematical model for copper homeostasis in *Enterococcus hirae*. Math Biosci **203**(2): 222-239.

Pedersen, R., Andersen, A.D., Molbak, L., Stagsted, J., and Boye, M. 2013. Changes in the gut microbiota of cloned and non-cloned control pigs during development of obesity: gut microbiota during development of obesity in cloned pigs. BMC Microbiol **13**: 30.

Pendle, S., Jelfs, P., Olma, T., Su, Y., Gilroy, N., and Gilbert, G.L. 2008. Difficulties in detection and identification of *Enterococcus faecium* with low-level inducible resistance to vancomycin, during a hospital outbreak. Clin Microbiol Infect **14**(9): 853-857.

Pesaro, M., Nicollier, G., Zeyer, J., and Widmer, F. 2004. Impact of soil drying-rewetting stress on microbial communities and activities and on degradation of two crop protection products. Appl Environ Microbiol **70**(5): 2577-2587.

Pfeiler, E.A., and Klaenhammer, T.R. 2007. The genomics of lactic acid bacteria. Trends Microbiol **15**(12): 546-553.

Picard, F.J., et al. 2004. Use of tuf sequences for genus-specific PCR detection and phylogenetic analysis of 28 streptococcal species. J Clin Microbiol **42**(8): 3686-3695.

Pompei, R., Berlutti, F., Thaller, M.C., Ingianni, A., Cortis, G., and Dainelli, B. 1992. *Enterococcus flavescens* sp. nov., a new species of enterococci of clinical origin. Int J Syst Bacteriol **42**(3): 365-369.

Pourcel, C., Salvignol, G., and Vergnaud, G. 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology **151**(Pt 3): 653-663.

Poyart, C., Quesnes, G., and Trieu-Cuot, P. 2000. Sequencing the gene encoding manganese-dependent superoxide dismutase for rapid species identification of enterococci. J Clin Microbiol **38**(1): 415-418.

Pride, D.T., et al. 2011. Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. Genome Res **21**(1): 126-136.

Priha, O., et al. 2013. Application of denaturing high-performance liquid chromatography for monitoring sulfate-reducing bacteria in oil fields. Appl Environ Microbiol **79**(17): 5186-5196.

Pryde, S.E., Richardson, A.J., Stewart, C.S., and Flint, H.J. 1999. Molecular analysis of the microbial diversity present in the colonic wall, colonic lumen, and cecal lumen of a pig. Appl Environ Microbiol **65**(12): 5372-5377.

Qin, J., et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature **464**(7285): 59-65.

Quednau, M., Ahrne, S., Petersson, A.C., and Molin, G. 1998. Identification of clinically important species of *Enterococcus* within 1 day with randomly amplified polymorphic DNA (RAPD). Curr Microbiol **36**(6): 332-336.

Rahkila, R., Johansson, P., Sade, E., and Bjorkroth, J. 2011. Identification of enterococci from broiler products and a broiler processing plant and description of *Enterococcus viikkiensis* sp. nov. Appl Environ Microbiol **77**(4): 1196-1203.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. 2004. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. Cell **118**(2): 229-241.

Ramotar, K., Conly, J.M., Chubb, H., and Louie, T.J. 1984. Production of menaquinones by intestinal anaerobes. J Infect Dis **150**(2): 213-218.

Ranjard, L., Lignier, L., and Chaussod, R. 2006. Cumulative effects of short-term polymetal contamination on soil bacterial community structure. Appl Environ Microbiol **72**(2): 1684-1687.

Reynolds, L.H., Packer, A., Bever, D.J., and Clay, K. 2003. Grassroots Ecology: Plant-Microbe-Soil Interactions as Drivers of Plant Community Structure and Dynamics. Ecology **84**(9): 2281-2291.

Rho, M., Wu, Y.W., Tang, H., Doak, T.G., and Ye, Y. 2012. Diverse CRISPRs evolving in human microbiomes. PLoS Genet **8**(6): e1002441.

Richter, M., and Rossello-Mora, R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A **106**(45): 19126-19131.

Robinson, I.M., Allison, M.J., and Bucklin, J.A. 1981. Characterization of the cecal bacteria of normal pigs. Appl Environ Microbiol **41**(4): 950-955.

Rodrigues, U., and Collins, M.D. 1990. Phylogenetic analysis of *Streptococcus saccharolyticus* based on 16S rRNA sequencing. FEMS Microbiol Lett **59**(1-2): 231-234.

Romani, A.M., Borrego, C.M., Diaz-Villanueva, V., Freixa, A., Gich, F., and Ylla, I. 2014. Shifts in microbial community structure and function in light- and dark-grown biofilms driven by warming. Environ Microbiol **16**(8): 2550-2567.

Round, J.L., and Mazmanian, S.K. 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol **9**(5): 313-323.

Ruan, J. 2013. [Bergey's Manual of Systematic Bacteriology (second edition) Volume 5 and the study of Actinomycetes systematic in China]. Wei Sheng Wu Xue Bao **53**(6): 521-530.

Ruiz-Garbajosa, P., et al. 2006. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. J Clin Microbiol **44**(6): 2220-2228.

Ryu, H., et al. 2013. Development of quantitative PCR assays targeting the 16S rRNA genes of *Enterococcus* spp. and their application to the identification of *Enterococcus* species in environmental samples. Appl Environ Microbiol **79**(1): 196-204.

Sakamoto, M., and Ohkuma, M. 2010. Usefulness of the hsp60 gene for the identification and classification of Gram-negative anaerobic rods. J Med Microbiol **59**(Pt 11): 1293-1302.

Savage, D.C. 2001. Microbial biota of the human intestine: a tribute to some pioneering scientists. Curr Issues Intest Microbiol **2**(1): 1-15.

Schellenberg, J., et al. 2009. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. Appl Environ Microbiol **75**(9): 2889-2898.

Schellenberg, J., Links, M.G., Hill, J.E., Hemmingsen, S.M., Peters, G.A., and Dumonceaux, T.J. 2011a. Pyrosequencing of chaperonin-60 (cpn60) amplicons as a means of determining microbial community composition. Methods Mol Biol **733**: 143-158.

Schellenberg, J.J., et al. 2011b. Molecular definition of vaginal microbiota in East African commercial sex workers. Appl Environ Microbiol **77**(12): 4066-4074.

Schleifer, K.H., and Kilpper-Balz, R. 1984. Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. Int. J. Syst. Bacteriol **34**: 31-34.

Schloss, P.D., Gevers, D., and Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One **6**(12): e27310.

Sedlacek, I., et al. 2013. *Enterococcus ureilyticus* sp. nov. and *Enterococcus rotai* sp. nov., two urease-producing enterococci from the environment. Int J Syst Evol Microbiol **63**(Pt 2): 502-510.

Sekirov, I., Russell, S.L., Antunes, L.C., and Finlay, B.B. 2010. Gut microbiota in health and disease. Physiol Rev **90**(3): 859-904.

Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A., and Delwart, E. 2011. The fecal virome of pigs on a high-density farm. J Virol **85**(22): 11697-11708.

Shapiro, B.J., et al. 2012. Population genomics of early events in the ecological differentiation of bacteria. Science **336**(6077): 48-51.

Shapiro, B.J., and Polz, M.F. 2014. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol **22**(5): 235-247.

Sherman. 1937. The Streptococci. Bacteriol Rev **1**(1): 3-97.

Sherman, J.M., and Wing, H.U. 1935. An unnoted hemolytic *Streptococcus* associated with milk products. . J. Dairy Sci **18(10)**: 657–660.

Simpson, P.J., Stanton, C., Fitzgerald, G.F., and Ross, R.P. 2003. Genomic diversity and relatedness of *bifidobacteria* isolated from a porcine cecum. J Bacteriol **185**(8): 2571-2581.

Sistek, V., et al. 2012. *Enterococcus ureasiticus* sp. nov. and *Enterococcus quebecensis* sp. nov., isolated from water. Int J Syst Evol Microbiol **62**(Pt 6): 1314-1320.

Slanetz, L.W., and Bartley, C.H. 1957. Numbers of enterococci in water, sewage, and feces determined by the membrane filter technique with an improved medium. J Bacteriol **74**(5): 591-595.

Sogin, M.L., et al. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A **103**(32): 12115-12120.

Solheim, M., Aakra, A., Snipen, L.G., Brede, D.A., and Nes, I.F. 2009. Comparative genomics of *Enterococcus faecalis* from healthy Norwegian infants. BMC Genomics **10**: 194.

Solioz, M., and Davies, K. 1994. Operon of vacuolar-type Na(+)-ATPase of *Enterococcus hirae*. J Biol Chem **269**(13): 9453-9459.

Solioz, M., and Stoyanov, J.V. 2003. Copper homeostasis in *Enterococcus hirae*. FEMS Microbiol Rev **27**(2-3): 183-195.

Sorek, R., Kunin, V., and Hugenholtz, P. 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol **6**(3): 181-186.

Stackebrandt, E., and Goebel, B.M. 1994. A place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol **44**: 846–849.

Stadtman, T.C. 1996. Selenocysteine. Annu Rev Biochem **65**: 83-100.

Staley, J.T., and Konopka, A. 1985. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu Rev Microbiol **39**: 321-346.

Struelens, M.J. 1998. Molecular epidemiologic typing systems of bacterial pathogens: current issues and perspectives. Mem Inst Oswaldo Cruz **93**(5): 581-585.

Sukontasing, S., Tanasupawat, S., Moonmangmee, S., Lee, J.S., and Suzuki, K. 2007. *Enterococcus camelliae* sp. nov., isolated from fermented tea leaves in Thailand. Int J Syst Evol Microbiol **57**(Pt 9): 2151-2154.

Suzzi, G., et al. 2000. A survey of the enterococci isolated from an artisanal Italian goat's cheese (semicotto caprino). J Appl Microbiol **89**(2): 267-274.

Svec, P., et al. 2001. *Enterococcus haemoperoxidus* sp. nov. and *Enterococcus moraviensis* sp. nov., isolated from water. Int J Syst Evol Microbiol **51**(Pt 4): 1567-1574.

Svec, P., and Sedlacek, I. 1999. Occurrence of *Enterococcus* spp. in waters. Folia Microbiol (Praha) **44**(1): 3-10.

Svec, P., et al. 2005a. *Enterococcus aquimarinus* sp. nov., isolated from sea water. Int J Syst Evol Microbiol **55**(Pt 5): 2183-2187.

Svec, P., et al. 2005b. *Enterococcus devriesei* sp. nov., associated with animal sources. Int J Syst Evol Microbiol **55**(Pt 6): 2479-2484.

Svec, P., et al. 2006. *Enterococcus silesiacus* sp. nov. and *Enterococcus termitis* sp. nov. Int J Syst Evol Microbiol **56**(Pt 3): 577-581.

Svec, P., et al. 2005c. Evaluation of (GTG)5-PCR for identification of *Enterococcus* spp. FEMS Microbiol Lett **247**(1): 59-63.

Svec, P., et al. 2012. *Enterococcus plantarum* sp. nov., isolated from plants. Int J Syst Evol Microbiol **62**(Pt 7): 1499-1505.

Swords, W.E., Wu, C.C., Champlin, F.R., and Buddington, R.K. 1993. Postnatal changes in selected bacterial groups of the pig colonic microflora. Biol Neonate **63**(3): 191-200.

Szabo, I., et al. 2009. Influence of a probiotic strain of *Enterococcus faecium* on *Salmonella enterica* serovar *Typhimurium* DT104 infection in a porcine animal infection model. Appl Environ Microbiol **75**(9): 2621-2628.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30**(12): 2725-2729.

Tan, C.K., et al. 2010. Bacteremia caused by non-faecalis and non-faecium *Enterococcus* species at a Medical center in Taiwan, 2000 to 2008. J Infect **61**(1): 34-43.

Tanasupawat, S., Sukontasing, S., and Lee, J.S. 2008. *Enterococcus thailandicus* sp. nov., isolated from fermented sausage ('mum') in Thailand. Int J Syst Evol Microbiol **58**(Pt 7): 1630-1634.

Tatusov, R.L., et al. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res **29**(1): 22-28.

Teixeira, L.M., et al. 2001. *Enterococcus porcinus* sp. nov. and *Enterococcus ratti* sp. nov., associated with enteric disorders in animals. Int J Syst Evol Microbiol **51**(Pt 5): 1737-1743.

Teixeira, L.M., Carvalho, M.G., Merquior, V.L., Steigerwalt, A.G., Brenner, D.J., and Facklam, R.R. 1997a. Phenotypic and genotypic characterization of *Vagococcus fluvialis*, including strains isolated from human sources. J Clin Microbiol **35**(11): 2778-2781.

Teixeira, L.M., et al. 1997b. Recent approaches on the taxonomy of the enterococci and some related microorganisms. Adv Exp Med Biol **418**: 397-400.

Teixeira, L.M., et al. 1996. Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and confirmation of *L. garvieae* as a senior subjective synonym of *Enterococcus seriolicida*. Int J Syst Bacteriol **46**(3): 664-668.

Tenover, F.C., Arbeit, R.D., and Goering, R.V. 1997. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. Infect Control Hosp Epidemiol **18**(6): 426-439.

Thiercelin, M. 1899. Morphologie et modes de reproduction de l'enterocoque. C R Seances Soc Biol Fil **11:551–553**.

Thompson, J. 1979. Lactose metabolism in *Streptococcus lactis*: phosphorylation of galactose and glucose moieties in vivo. J Bacteriol **140**(3): 774-785.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22**(22): 4673-4680.

Torok, V.A., Allison, G.E., Percy, N.J., Ophel-Keller, K., and Hughes, R.J. 2011. Influence of antimicrobial feed additives on broiler commensal posthatch gut microbiota development and performance. Appl Environ Microbiol **77**(10): 3380-3390.

Torsvik, V., Goksoyr, J., and Daae, F.L. 1990. High diversity in DNA of soil bacteria. Appl Environ Microbiol **56**(3): 782-787.

Torsvik, V., Ovreas, L., and Thingstad, T.F. 2002. Prokaryotic diversity--magnitude, dynamics, and controlling factors. Science **296**(5570): 1064-1066.

Tran, H., et al. 2012. Effects of lactose and yeast-dried milk on growth performance, fecal microbiota, and immune parameters of nursery pigs. J Anim Sci **90**(9): 3049-3059.

Travisano, M. 1997. Long-term experimental evolution in *Escherichia coli*. VI. Environmental constraints on adaptation and divergence. Genetics **146**(2): 471-479.

Tsiodras, S., Gold, H.S., Coakley, E.P., Wennersten, C., Moellering, R.C., Jr., and Eliopoulos, G.M. 2000. Diversity of domain V of 23S rRNA gene sequence in different *Enterococcus* species. J Clin Microbiol **38**(11): 3991-3993.

Tyrrell, G.J., Bethune, R.N., Willey, B., and Low, D.E. 1997. Species identification of enterococci via intergenic ribosomal PCR. J Clin Microbiol **35**(5): 1054-1060.

Tyrrell, G.J., et al. 2002. *Enterococcus gilvus* sp. nov. and *Enterococcus pallens* sp. nov. isolated from human clinical specimens. J Clin Microbiol **40**(4): 1140-1145.

Tyson, G.W., et al. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428**(6978): 37-43.

Valdezate, S., et al. 2007. Epidemiological tracing of *Salmonella enterica* serotype Abortusovis from Spanish ovine flocks by PFGE fingerprinting. Epidemiol Infect **135**(4): 695-702.

van Schaik, W., et al. 2010. Pyrosequencing-based comparative genome analysis of the nosocomial pathogen *Enterococcus faecium* and identification of a large transferable pathogenicity island. BMC Genomics **11**: 239.

Vancanneyt, M., et al. 2001. Enterococcus villorum sp. nov., an enteroadherent bacterium associated with diarrhoea in piglets. Int J Syst Evol Microbiol **51**(Pt 2): 393-400.

Vancanneyt, M., et al. 2004. *Enterococcus saccharominimus* sp. nov., from dairy products. Int J Syst Evol Microbiol **54**(Pt 6): 2175-2179.

Venter, J.C., et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**(5667): 66-74.

Verbeke, T.J., Sparling, R., Hill, J.E., Links, M.G., Levin, D., and Dumonceaux, T.J. 2011. Predicting relatedness of bacterial genomes using the chaperonin-60 universal target (cpn60 UT): application to *Thermoanaerobacter* species. Syst Appl Microbiol **34**(3): 171-179.

Vermette, C.J., Russell, A.H., Desai, A.R., and Hill, J.E. 2010. Resolution of phenotypically distinct strains of *Enterococcus* spp. in a complex microbial community using cpn60 universal target sequencing. Microb Ecol **59**(1): 14-24.

Versalovic, J., Koeuth, T., and Lupski, J.R. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res **19**(24): 6823-6831.

Vorontsova, M.S., Clayton, D., and Simon, B.K. 2015. Grassroots e-floras in the Poaceae: growing GrassBase and GrassWorld. PhytoKeys(48): 73-84.

Vos, M., Quince, C., Pijl, A.S., de Hollander, M., and Kowalchuk, G.A. 2012. A comparison of rpoB and 16S rRNA as markers in pyrosequencing studies of bacterial diversity. PLoS One **7**(2): e30600.

Wagner, M., Amann, R., Lemmer, H., and Schleifer, K.H. 1993. Probing activated sludge with oligonucleotides specific for proteobacteria: inadequacy of culture-dependent methods for describing microbial community structure. Appl Environ Microbiol **59**(5): 1520-1525.

Ward, D.M., Weller, R., and Bateson, M.M. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature **345**(6270): 63-65.

Watabe, M., et al. 2008. Epidemiological study of *E. coli* O157:H7 isolated in Northern Ireland using pulsed-field gel electrophoresis (PFGE). Ulster Med J **77**(3): 168-174.

Wattam, A.R., et al. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res **42**(Database issue): D581-591.

Weng, P.L., Ramli, R., Shamsudin, M.N., Cheah, Y.K., and Hamat, R.A. 2013. High genetic diversity of *Enterococcus faecium* and *Enterococcus faecalis* clinical isolates by pulsed-field gel electrophoresis and multilocus sequence typing from a hospital in Malaysia. Biomed Res Int **2013**: 938937.

Wiedenbeck, J., and Cohan, F.M. 2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev **35**(5): 957-976.

Wieser, M., and Busse, H.J. 2000. Rapid identification of *Staphylococcus epidermidis*. Int J Syst Evol Microbiol **50 Pt 3**: 1087-1093.

Willems, R.J., et al. 2005. Global spread of vancomycin-resistant *Enterococcus faecium* from distinct nosocomial genetic complex. Emerg Infect Dis **11**(6): 821-828.

Willems, R.J., et al. 2012. Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. MBio **3**(4): e00151-00112.

Williams, A.M., Farrow, J.A.E., and Collins, M.D. 1989. Reverse transcriptase sequencing of 16S ribosomal RNA from *Streptococcus cecorum*. Lett Appl Microbiol **8**(5): 185-189.

Williams, A.M., Rodrigues, U.M., and Collins, M.D. 1991. Intrageneric relationships of Enterococci as determined by reverse transcriptase sequencing of small-subunit rRNA. Res Microbiol **142**(1): 67-74.

Winslow, C.E. 1914. The Characterization and Classification of Bacterial Types. Science **39**(994): 77-91.

Woese, C.R., and Fox, G.E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A **74**(11): 5088-5090.

Wu, G.D., et al. 2011a. Linking long-term dietary patterns with gut microbial enterotypes. Science **334**(6052): 105-108.

Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., and van der Lelie, D. 2011b. Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. FEMS Microbiol Rev **35(2)**: 299-323.

Wunderli-Ye, H., and Solioz, M. 1999. Copper homeostasis in *Enterococcus hirae*. Adv Exp Med Biol **448**: 255-264.

Yamaguchi, H., et al. 1999. Induction of secretion of interleukin-8 from human gastric epithelial cells by heat-shock protein 60 homologue of *Helicobacter pylori*. J Med Microbiol **48**(10): 927-933.

Yamaguchi, H., et al. 1997. Heat-shock protein 60 homologue of *Helicobacter pylori* is associated with adhesion of H. pylori to human gastric epithelial cells. J Med Microbiol **46**(10): 825-831.

Yolton, D.P., and Savage, D.C. 1976. Influence of certain indigenous gastrointestinal microorganisms on duodenal alkaline phosphatase in mice. Appl Environ Microbiol **31**(6): 880-888.

Yousif, N.M., et al. 2005. Molecular characterization, technological properties and safety aspects of enterococci from 'Hussuwa', an African fermented sorghum product. J Appl Microbiol **98**(1): 216-228.

Zankari, E., et al. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **67**(11): 2640-2644.

Zhang, Y., and Gladyshev, V.N. 2008. Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project. PLoS Genet **4**(6): e1000095.

Zhang, Y., Romero, H., Salinas, G., and Gladyshev, V.N. 2006. Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. Genome Biol **7**(10): R94.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. 2011. PHAST: a fast phage search tool. Nucleic Acids Res **39**(Web Server issue): W347-352.

Zuniga, M., et al. 2005. Horizontal gene transfer in the molecular evolution of mannose PTS transporters. Mol Biol Evol **22**(8): 1673-1685.